

From: Schnizer, Holly
Sent: Tuesday, July 03, 2001 7:47 AM
To: STIC-Biotech/ChemLib
Subject: sequence search for appl. no. 09/490,291

Please search the commercial and interference databases for the following:

SEQ ID NO: 2, 4, 6, 8, 9 and 11

Thank you.

Holly Schnizer
AU 1653
CM1-10B05
305-3722

RELATED
PCT 00/30086

qBel
MB

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 12C14 Tel: 308-44984

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:14 ; Search time 86.32 Seconds
(without alignments)
51.971 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377
Sequence: 1 MRSHHHHHGMASGDLKN.....AKIGDLNNTSGIRPPAAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq.0601.*
1: /SID88/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	279.5	74.1	76 21	AA10562
2	243.5	64.6	21 21	AA10561
3	221	58.6	84 19	AAW43025
4	213	56.5	84 19	AAW43026
5	205	54.4	84 19	AAW43027
6	199	52.8	42 21	AA10558
7	197	52.3	84 19	AAW43028
8	181	48.0	84 19	AAW43024
9	159	42.2	42 21	AA10557
10	124	32.9	42 19	AAW43029
11	121	32.1	43 21	AA108364

12	121	32.1	43	21	AA108383
13	119.5	31.7	103	22	AA107097
14	118.5	31.4	93	22	AA107096
15	115	30.5	43	21	AA108344
16	114.5	30.4	43	21	AA108375
17	113	30.0	43	21	AA108362
18	113	30.0	43	21	AA108365
19	111.5	29.6	43	21	AA108386
20	110.4	27.6	43	21	AA108378
21	101.5	26.9	286	21	AA156997
22	101.5	26.9	286	22	AA156978
23	96.5	25.6	491	18	AA13373
24	96	25.5	175	21	AA144383
25	95	25.2	373	14	AA139486
26	94	24.9	168	21	AA144384
27	94	24.9	169	21	AA144380
28	94	24.9	175	21	AA144382
29	92	24.4	145	20	AA194266
30	92	24.4	159	19	AA131552
31	92	24.4	493	12	AA13982
32	91.5	24.3	168	21	AA144385
33	91.5	24.3	169	21	AA144381
34	91	24.1	18	21	AA10560
35	91	24.1	379	10	AA190508
36	90.5	24.0	380	10	AA190510
37	90.5	24.0	380	10	AA190511
38	90.5	24.0	380	10	AA190512
39	90.5	24.0	337	14	AA139485
40	90	23.9	337	14	AA139485
41	90	23.9	345	19	AA131555
42	90	23.9	359	19	AA161046
43	89.5	23.7	342	14	AA139487
44	89.5	23.7	342	14	AA139489
45	89.5	23.7	342	14	AA139491

ALIGNMENTS

RESULT 1
ID AAB10562 standard: protein; 76 AA.
XX AAB10562;
AC AAB10562;
XX 15-DEC-2000 (first entry)
DT
XX
DE Copolymer block protein insert L2-B.
XX
KW Coiled coil protein; copolymer blocker; vulnery; wound treatment;
KW integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KW antibiotic; dressing; abrasion; burn.
XX
OS Unidentified.
XX
PN US6090911-A.
XX
PD 18-JUL-2000.
XX
PF 22-OCT-1997; 97US-0956307.
XX
PR 22-OCT-1997; 97US-0956307.
XX
PA (UYMA-) UNIV MASSACHUSETTS.
XX
PI Petka WA, McGrath KP, Tirrell DA.
XX
XX WPI; 2000-542235/49.
XX
XX Synthetic block copolymer with an antibiotic compound, useful for
XX dressing abrasion, burn or non-puncture wound, comprises two
XX alpha-helical protein blocks, random-coil protein block and linker
XX proteins -

Peptide used to as
HIV protease-associ
HIV protease-associ
Amino acid sequenc
Immoblilised subste
Amino acid sequenc
Amino acid sequenc
Peptide used to as
ZLP5 polypeptide u
Human procaspase 8
Human procaspase-8.
Recombinant Sp 100
P. aeruginosa E coli
Human apoAIV mutel
P. aeruginosa H co
P. aeruginosa H co
P. aeruginosa H co
H6PXTIPB-UB fusio
Collagen binding P
P. falcatiparum spor
P. aeruginosa E coli
P. aeruginosa E coli
Coiled coil protei
Sequence of an epi
Sequence of an epi
Sequence of an epi
Sequence of an epi
Sequence of new po
Human apoAIV mutel
Fibronectin-bindin
His-tagged polylin
Human apoAIV mutel
Human apoAIV mutel
Human apoAIV mutel

DR WPI: 1998-120029/11.
 XX
 PT Production of peptide hetero:dimer(s) - in the fabrication of
 XX self-assembling protein-based structural material(s)
 XX
 XX
 XX Disclosure: Fig 4B; 17pp; English.
 CC
 CC AAM43024-28 represent recognition sequences obtained from the library of
 CC recognition sequences coded for by AAV04943. This DNA sequence is
 CC obtained by reverse translation. The codons were chosen to maximise
 CC expression in *Escherichia coli*, and to introduce useful restriction for
 CC subsequent genetic manipulations. The DNA sequence use a "mixed site"
 CC approach at the first base of the codons for amino acids positions 5 and
 CC 7 of the heptad (see below). The formation of a self-assembling
 CC structural material is mediated by the artificial recognition sequences
 CC (e.g. present sequence). A method of producing a self-assembling
 CC protein-based structural material comprises mixing a multiplicity of
 CC 2 different artificial peptide sequences, designed so as to
 CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil
 CC alpha helical secondary structure having at least 2 heptad units. The
 CC artificial peptides spontaneously heterodimerise into a self-assembled
 CC protein-based structural material. The method is used to produce
 CC nanoscale structural materials designed for specific functions e.g.
 CC membranes, fibres, absorbents, reactive materials, etc. A specific
 CC application is in the fabrication of selectively permeable membranes for
 CC protective gloves and coated fabrics for fuel handlers and other
 CC hazardous materials. Materials for medical implants, wound healants and
 CC other medical treatments would also be feasible using this technology.
 XX
 XX
 XX Sequence 84 AA:

```

OY      16  GDLKNKVAQLKRRVSLKDKAELKQEVSRLENIEDLAKTIGDLNN 62
          |||||||
Db      2  gdlknkvaglkrrkvrslkdkaaelkqevsrleiedlkkkigdlkn 48
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0

RESULT  4
AAW43026
ID      AAW43026 standard; Peptide: 84 AA.
XX
XX      AAW43026;
XX
XX      08-MAY-1998 (first entry)
XX
XX      Artificial recognition sequence 3.
XX
XX      Recognition sequence: self-assembling; protein-based structural material;
KW      spontaneous heterodimerisation; coiled-coil alpha helix;
KW      selectively permeable membrane; coated fabric.
XX
XX      Synthetic.
XX
XX      US5712366-A.
XX
XX      27-JAN-1998.
XX
XX      25-MAY-1995; 95US-0452592.
XX
XX      25-MAY-1993; 93US-0068948.
XX
XX      (USSA ) US SEC OF ARMY.
XX
XX      Kaplan DL, McGrath KP;
XX
XX      WPI; 1998-120029/11.
XX
XX      Production of peptide hetero:dimer(s) - in the fabrication of
XX      self-assembling protein-based structural material(s)

```

PS Disclosure; Fig 4C; 17pp; English.

XX
XX
CC AMN43024-28 represent recognition sequences obtained from the library of
CC recognition sequences coded for by AAV04943. This DNA sequence is
CC obtained by reverse translation. The codons were chosen to maximise
CC expression in Escherichia coli, and to introduce useful restriction for
CC subsequent genetic manipulations. The DNA sequence uses a "mixed site"
CC approach at the first base of the codons for amino acid positions 5 and
CC 7 of the heptad (see below). The formation of a self-assembling
CC structural material is mediated by the artificial recognition sequences
CC (e.g. present sequence). A method of producing a self-assembling
CC protein-based structural material comprises mixing a multiplicity of
CC 2 different artificial peptide sequences, designed so as to
CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil
CC alpha helical secondary structure having at least 2 heptad units. The
CC artificial peptides spontaneously heterodimerise into a self-assembled
CC protein-based structural material. The method is used to produce
CC nanoscale structural materials designed for specific functions e.g.
CC membranes, fibres, absorbents, reactive materials, etc. A specific
CC application is in the fabrication of selectively permeable membranes for
CC protective gloves and coated fabrics for fuel handlers and other
CC hazardous materials. Materials for medical implants, wound healants and
CC other medical treatments would also be feasible using this technology.

XX
XQ Sequence 84 AA:

```

Query Match 56.5%; Score 213; DB 19; Length 84;
Best Local Similarity 93.6%; Pred. No. 1.2e-15;
Matches 44; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKNKVAQLRRKRVSLKDKAELKQEVSRLENETEDLKAKIGDINN 62
      |||||:::|||||
Db 2 gdlknkvaglerevrsldkkaaelkqevsrleenedlkakigdlkn 48

RESULT 5
AAW43027
ID AAW43027 standard; Peptide: 84 AA.
AC AAW43027;
XX AAW43027;
XX 08-MAY-1998 (first entry)
XX
XX Artificial recognition sequence 4.
DE
XX Recognition sequence; self-assembling; protein-based structural material;
KW spontaneous heterodimerisation; coiled-coil alpha helix;
KW selectively permeable membrane; coated fabric.
XX
XX Synthetic.
OS
XX US5712366-A.
XX
XX 27-JAN-1998.
XX
XX 25-MAY-1995; 95US-0452592.
XX
XX 25-MAY-1993; 93US-0068948.
XX
XX (USSA ) US SEC OF ARMY.
XX
XX Kaplan DL, McGrath KP;
XX
XX WPI; 1998-120029/11.
XX
XX
XX Production of peptide hetero:dimer(s) - in the fabrication of
XX self-assembling protein-based structural material(s)
XX
XX Disclosure; Fig 4D; 17pp; English.
XX
XX AAW43024-28 represent recognition sequences obtained from the library of

```


CC 7 of the heptad (see below). The formation of a self-assembling
CC structural material is mediated by the artificial recognition sequences
CC (e.g. present sequence). A method of producing a self-assembling
CC protein-based structural material comprises mixing a multiplicity of
CC 2 different artificial peptide sequences, designed so as to
CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil
CC alpha helical secondary structure having at least 2 heptad units. The
CC artificial peptides spontaneously heterodimerise into a self-assembled
CC protein-based structural material. The method is used to produce
CC nanoscale structural materials designed for specific functions e.g.
CC membranes, fibres, absorbants, reactive materials, etc. A specific
CC application is in the fabrication of selectively permeable membranes for
CC protective gloves and coated fabrics for fuel handlers and other
CC hazardous materials. Materials for medical implants, wound healants and
CC other medical treatments would also be feasible using this technology.

XX Sequence 84 AA:

SO

Query Match 52.3%; Score 197; DB 19; Length 84;
Best Local Similarity 85.1%; Pred. No. 6e-14; 1; Indels 0; Gaps 0;
Matches 40; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDKKNVAVQLRRKRVSLKDKAELEKQEVSRLENEIEDLKAKIGDLNN 62
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 2 gdlenevaqlerevrsledeaaelkqevsrleeneiedlkakigdlenn 48

RESULT 8
AAM43024
ID AAM43024 standard; Peptide; 84 AA.
XX
AC AAM43024;
XX
DT 08-MAY-1998 (first entry)
XX
DE Artificial recognition sequence 1.
XX
KM Recognition sequence: self-assembling; protein-based structural material;
KM spontaneous heterodimerisation; coiled-coil alpha helix;
KM selectively permeable membrane; coated fabric.
XX
OS Synthetic.
XX
PN US5712366-A.
XX
PD 27-JAN-1998.
XX
PF 25-MAY-1995; 95US-0452592.
XX
PR 25-MAY-1993; 93US-0068948.
XX
PA (USSA) US SEC OF ARMY.
XX
PI Kaplan DL, McGrath KP;
XX
DR WPI, 1998-120029/11.
XX
PT Production of peptide hetero:dimer(s) - in the fabrication of
PT self-assembling protein-based structural material(s)
XX
PS Disclosure; Fig 4A; 17pp; English.
XX
CC AAM43024-28 represent recognition sequences obtained from the library of
CC recognition sequences coded for by AAV04943. This DNA sequence is
CC obtained by reverse translation. The codons were chosen to maximise
CC expression in *Escherichia coli*, and to introduce useful restriction for
CC subsequent genetic manipulations. The DNA sequence use a "mixed site"
CC approach at the first base of the codons for amino acids positions 5 and
CC 7 of the heptad (see below). The formation of a self-assembling
CC structural material is mediated by the artificial recognition sequences
CC (e.g. present sequence). A method of producing a self-assembling
CC protein-based structural material comprises mixing a multiplicity of

CC 2 different artificial peptide sequences, designed so as to
CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil
CC alpha helical secondary structure having at least 2 heptad units. The
CC artificial peptides spontaneously heterodimerise into a self-assembled
CC protein-based structural material. The method is used to produce
CC nanoscale structural materials designed for specific functions e.g.
CC membranes, fibres, absorbants, reactive materials, etc. A specific
CC application is in the fabrication of selectively permeable membranes for
CC protective gloves and coated fabrics for fuel handlers and other
CC hazardous materials. Materials for medical implants, wound healants and
CC other medical treatments would also be feasible using this technology.

XX Sequence 84 AA:

SO

Query Match 48.0%; Score 181; DB 19; Length 84;
Best Local Similarity 76.6%; Pred. No. 3e-12; 1; Indels 0; Gaps 0;
Matches 36; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDKKNVAVQLRRKRVSLKDKAELEKQEVSRLENEIEDLKAKIGDLNN 62
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 2 gdlenevaqlerevrsledeaaelkqevsrleeneiedlkakigdlenn 48

RESULT 9
AAB10557
ID AAB10557 standard; protein; 42 AA.
XX
AC AAB10557;
XX
DT 15-DEC-2000 (first entry)
XX
DE Coiled coil protein A1.
XX
KM Coiled coil protein; copolymer blocker; vulnerray; wound treatment;
KM integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KM antibiotic; dressing; abrasion; burn; A1.
XX
OS Unidentified.
XX
PN US6090911-A.
XX
PD 18-JUL-2000.
XX
PF 22-OCT-1997; 97US-0956307.
XX
PR 22-OCT-1997; 97US-0956307.
XX
PA (UYMA-) UNIV MASSACHUSETTS.
XX
PI Petka WA, McGrath KP, Tirrell DA;
XX
DR WPI, 2000-542235/49.
XX
PT Synthetic block copolymer with an antibiotic compound, useful for
PT dressing abrasion, burn or non-puncture wound, comprises two
PT alpha-helical protein blocks, random-coil protein block and linker
PT proteins -
XX
PS Example 1; Column 25-26; 53pp; English.
XX
CC This invention describes a novel synthetic block copolymer XYZ comprising
CC two alpha-helical protein blocks X and Z, a water soluble, random-coil
CC protein block Y which links X and Z, and linker proteins for linking the
CC alpha-helical protein blocks to the random-coil protein block. The
CC products of the invention have vulnerray activity. A copolymer produced
CC from a gene that encodes a derivative of AC_1_0A copolymer was suspended
CC in water to form a gel, which was then used to treat a wound. The cell
CC binding domain was the integrin ArgGlyAspser sequence which is known to
CC bind to gp130/IL6 proteins expressed on fibroblast cells necessary for
CC matrix formation for the regeneration of skin. Fibroblasts become
CC entrapped within the gel and thus remain at the site of the wound,
CC serving as a scaffold for the regeneration of tissue surrounding the

CC wound. A synthetic block copolymer together with an antibiotic compound
CC is useful for wound dressing. The copolymer is useful for dressing
CC abrasion, burn or non-puncture wound. Unlike conventional polymers, the
CC polymers of equal size are produced biologically from a single template.
CC The intermediate binding of alpha-helical blocks are monodisperse which
CC gives a uniform pore size depending on the length of random-coil block.
CC The new copolymers have low molecular weight hence have decreased
CC viscosity solutions or gels under suitable conditions. This sequence
CC represents the coiled coil protein A1 which is used to illustrate the
CC method of the invention.

XX Sequence 42 AA:

Query Match 42.2%; Score 159; DB 21; Length 42;

Best Local Similarity 76.2%; Pred. No. 2, 9e-10;
Matches 33; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY 15 SGDLKNKVAQLKRRKYSIKDKRAELKQVSRLENEDLKAK 56
Db 1 sgdlenevaqlerevrsledeaeleqvsrlnkneidlkae 42

RESULT 10

AAW43029 ID AAW43029 standard; Peptide: 42 AA.

XX AC AAW43029;

DT 08-MAY-1998 (first entry)

XX Amino acid sequence giving a library of recognition sequences.

XX Recognition sequence: self-assembling; protein-based structural material;
KW spontaneous heterodimerisation; coiled-coil alpha helix;
KM selectively permeable membrane; coated fabric.

XX Synthetic.

XX Key Location/Qualifiers

FT MISC-difference 5 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 7 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 12 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 14 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 19 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 21 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 26 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 28 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 33 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 35 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 40 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 42 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 42 /note= "any charged residue, preferably Glu or Lys"

XX US5712366-A.

XX 27-JAN-1998.

XX 25-MAY-1995; 95US-0452592.

XX 25-MAY-1993; 93US-0068948.

PA (USSA) US SEC OF ARMY.

PI Kaplan DL, McGrath KP;

XX WPI: 1998-120029/11.

DR N-PSDB; AAV04943.

PT Production of peptide hetero: dimer(s) - In the fabrication of
PT self-assembling protein-based structural material(s)

PS Claim 2; Columns 13-14; 17pp; English.

XX The present sequence is used to obtain an entire library of recognition
CC sequences. The DNA sequence encoding the present sequence is obtained by
CC reverse translation. The codons were chosen to maximise expression in
CC Escherichia coli, and to introduce useful restriction for subsequent
CC genetic manipulations. The DNA sequence use a "mixed site" approach at
CC the first base of the codons for amino acids positions 5 and 7 of the
CC heptad (see below). The formation of a self-assembling structural
CC material is mediated by the artificial recognition sequences obtained
CC from the present sequence. A method of producing a self-assembling
CC protein-based structural material comprises mixing a multiplicity of
CC 2 different artificial peptide sequences, designed so as to spontaneously
CC heterodimerise. Both artificial peptides have a coiled-coil alpha helical
CC secondary structure having at least 2 heptad units. The artificial
CC peptides spontaneously heterodimerise into a self-assembled protein-based
CC structural material. The method is used to produce nanoscale structural
CC materials designed for specific functions e.g. "membranes, fibres",
CC absorbants, reactive materials, etc. A specific application is in the
CC fabrication of selectively permeable membranes for protective gloves and
CC coated fabrics for fuel handlers and other hazardous materials. Materials
CC for medical implants, wound healants and other medical treatments would
CC also be feasible using this technology.

XX Sequence 42 AA:

Query Match 32.9%; Score 124; DB 19; Length 42;
Best Local Similarity 72.5%; Pred. No. 1, 4e-06;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 16 GDILKNKVAQLKRRKYSIKDKRAELKQVSRLENEDLKAK 55
Db 2 gdlxnvaqlrxrvrsldxaalxqvsrlnxkndlxa 41

RESULT 11

AAB08364 ID AAB08364 standard; Peptide: 43 AA.

XX AAB08364;

DT 20-DEC-2000 (first entry)

XX Amino acid sequence of a coiled-coil peptide.

XX Binding partner; protein modification; post-translational modification;
KW modulator; coiled-coil structure.

XX Unidentified.

XX Key Location/Qualifiers

FT Cleavage-site 24 /note= "thrombin cleavage site"

FT MISC-difference 39 /note= "fluorophore attachment site"

XX W0200050902-A2.

XX 31-AUG-2000.

XX 25-FEB-2000; 2000WO-GB00669.

PR 25-FEB-1999; 99GB-0004398.
XX (FLUO-) FLUORESCIENCE LTD.
XX Colyer J, Craig RK, Maschio A, Mezna M;
PI WPI; 2000-572119/53.
DR
XX
XX High throughput assay for monitoring modification of polypeptides and
PT modulation of the modifications
PS Disclosure; Page 61; 128pp; English.
XX
CC The specification describes a method for analysing a sample. The method
CC comprises immobilising a polypeptide to a physical support, contacting
CC the immobilised polypeptide with a test sample which may contain an
CC agent capable of modifying the immobilised polypeptide, contacting the
CC immobilised polypeptide with a binding partner polypeptide, where
CC association of both polypeptide is dependent on the modification
CC of the immobilised polypeptide, and measuring the association of
CC the binding partner polypeptide to the immobilised polypeptide. The
CC polypeptides, support and methods can be used to analyse a sample to
CC determine if modification of a polypeptide is taking place and to
CC identify modulators of the modification. This is useful for monitoring
CC the post-translational modification of proteins. AAB08364-65 represent
CC binding partners which have coiled-coil structures, and may be used in
CC the method of the invention to assay for thrombin activity.
XX
SQ Sequence 43 AA:

Query Match 32.1%; Score 121; DB 21; Length 43;
Best Local Similarity 46.7%; Pred. No. 3.1e-06;
Matches 21; Conservative 14; Mismatches 8; Indels 2; Gaps 1;

OY 5 HHHHHGSMAGDLKNVADLRKVRSLKDKAELKQVSRLENE 49
Db 1 hhhhhgsla--qlegeladqegenrqldeqeladqlegetakleage 43

RESULT 12
AAB08383
ID AAB08383 standard; peptide; 43 AA.
XX
AC AAB08383;
XX
DT 20-DEC-2000 (first entry)
XX
DE Peptide used to assay for tobacco etch virus protease activity.
XX
XX Binding partner; protein modification; post-translational modification;
KW modulator; coiled-coil structure.
XX
OS Unidentified.
XX
XX WO200050902-A2.
PN
XX
PD 31-AUG-2000.
XX
PF 25-FEB-2000; 2000WO-GB00669.
XX
XX 25-FEB-1999; 99GB-0004398.
PR
XX (FLUO-) FLUORESCIENCE LTD.
PA
XX Colyer J, Craig RK, Maschio A, Mezna M;
PI WPI; 2000-572119/53.
DR
XX High throughput assay for monitoring modification of polypeptides and
PT modulation of the modifications
PS Example 9; Page 80; 128pp; English.

XX
CC The specification describes a method for analysing a sample. The method
CC comprises immobilising a polypeptide to a physical support, contacting
CC the immobilised polypeptide with a test sample which may contain an
CC agent capable of modifying the immobilised polypeptide, contacting the
CC immobilised polypeptide with a binding partner polypeptide, where
CC association of both polypeptide is dependent on the modification
CC of the immobilised polypeptide, and measuring the association of
CC the binding partner polypeptide to the immobilised polypeptide. The
CC polypeptides, support and methods can be used to analyse a sample to
CC determine if modification of a polypeptide is taking place and to
CC identify modulators of the modification. This is useful for monitoring
CC the post-translational modification of proteins. AAB08383-84 represent
CC binding partners which are used, in the method of the invention, to
CC assay for tobacco etch virus (TEV) protease activity.
XX
SQ Sequence 43 AA:

Query Match 32.1%; Score 121; DB 21; Length 43;
Best Local Similarity 46.7%; Pred. No. 3.1e-06;
Matches 21; Conservative 14; Mismatches 8; Indels 2; Gaps 1;

OY 5 HHHHHGSMAGDLKNVADLRKVRSLKDKAELKQVSRLENE 49
Db 1 hhhhhgsla--qlegeladqegenrqldeqeladqlegetakleage 43

RESULT 13
AAB70797
ID AAB70797 standard; Protein; 103 AA.
XX
AC AAB70797;
XX
DT 25-MAY-2001 (first entry)
XX
DE HIV protease-associated protein #2.
XX
XX Protease substrate; detection; cleavage site; stretched helix;
KW screening.
XX
OS Human immunodeficiency virus.
XX
XX DE19941239-A1.
PN
XX
PD 01-MAR-2001.
XX
PF 26-AUG-1999; 99DE-1041239.
XX
XX 26-AUG-1999; 99DE-1041239.
PR
XX (IMB-) IMB INST MOLEKULARE BIOTECHNOLOGIE EV.
PA
XX
PI Aldinger U, Hill O, Hallsch A, Steinruecke P;
XX
XX WPI; 2001-246019/26.
DR
XX
XX Protease substrate useful for measuring protease activities and in
PT online screening assays for proteases or protease inhibitors, comprises
PT a polypeptide with a binding site, a cleavage site and a stretched
PT helix -
XX
XX Disclosure; Fig 2; 14pp; German.
PS
XX This invention describes a novel protease substrate for detecting a
CC target protease which comprises a polypeptide containing a high-affinity
CC binding site (1) for binding to a support, a cleavage site (2) specific
CC for the target protease, and a sequence of amino acids forming a
CC stretched helix (3). The substrate is useful for measuring protease
CC activities and in online screening assays for proteases or protease
CC inhibitors, especially based on surface plasmon resonance.
XX
SQ Sequence 103 AA:

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:46 ; Search time 48.42 Seconds
(without alignments)
30.787 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377
Sequence: 1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRRAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 segs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
- 4: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
- 5: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
- 6: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	279.5	74.1	76	3	US-08-956-307B-18
2	243.5	64.6	76	3	US-08-956-307B-17
3	221	58.6	84	1	US-08-452-592B-7
4	213	56.5	84	1	US-08-452-592B-8
5	205	54.4	84	1	US-08-452-592B-9
6	199	52.8	42	3	US-08-956-307B-2
7	197	52.3	84	1	US-08-452-592B-10
8	181	48.0	84	1	US-08-452-592B-6
9	159	42.2	42	3	US-08-956-307B-1
10	124	32.9	42	3	US-08-452-592B-11
11	91	24.1	18	3	US-08-956-307B-4
12	87	23.1	14	3	US-08-956-307B-3
13	82.5	21.9	110	1	US-08-434-705B-15
14	82.5	21.9	110	2	US-08-086-201-15
15	82	21.8	18	1	US-08-578-649-14
16	81	21.5	42	5	PCT-US94-06655-9
17	81	21.5	48	1	US-08-457-245-21
18	81	21.5	126	1	US-08-268-348A-12
19	81	21.5	133	1	US-08-268-348A-8
20	81	21.5	133	1	US-08-268-348A-10
21	79.5	21.1	31	2	US-08-662-227-37
22	79	21.0	32	3	US-08-737-336-1
23	79	21.0	39	2	US-08-679-865-39
24	79	21.0	39	2	US-08-680-876-39
25	79	21.0	69	2	US-08-687-865A-21
26	79	21.0	69	4	US-09-043-711-21
27	79	21.0	88	2	US-08-690-011A-11

28	79	21.0	254	4	US-09-004-731-89	Sequence 89, App1
29	79	21.0	254	4	US-09-004-731-92	Sequence 92, App1
30	79	21.0	254	4	US-08-749-699-89	Sequence 89, App1
31	79	21.0	254	4	US-08-749-699-92	Sequence 92, App1
32	79	21.0	351	2	US-08-701-191A-3	Sequence 3, App1
33	79	21.0	532	3	US-08-737-336-6	Sequence 6, App1
34	79	21.0	631	1	US-08-605-541B-12	Sequence 12, App1
35	78.5	20.8	98	2	US-08-690-011A-19	Sequence 19, App1
36	78	20.7	16	1	US-08-578-649-13	Sequence 13, App1
37	78	20.7	20	4	US-08-640-737-32	Sequence 32, App1
38	78	20.7	35	2	US-08-491-527A-2	Sequence 2, App1
39	78	20.7	35	2	US-08-845-659-2	Sequence 2, App1
40	78	20.7	35	3	US-08-845-658-2	Sequence 2, App1
41	78	20.7	122	3	US-08-956-307B-19	Sequence 19, App1
42	78	20.7	184	3	US-09-023-082A-30	Sequence 30, App1
43	75.5	20.0	153	4	US-08-676-318A-34	Sequence 34, App1
44	75	19.9	12	5	PCT-US94-09700-4	Sequence 4, App1
45	75	19.9	20	2	US-08-543-020-12	Sequence 12, App1

ALIGNMENTS

RESULT 1
US-08-956-307B-18
Sequence 18, Application US/08956307B
Patent No. 6090911
GENERAL INFORMATION:
APPLICANT: Petka, Wendy A.
APPLICANT: Tirrell, David A.
APPLICANT: Kevin P. McGrath
TITLE OF INVENTION: REVERSIBLE HYDROGELS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,307B
FILING DATE: 22-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07880/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-956-307B-18

Query Match 74.1%; Score 279.5; DB 3; Length 76;
Best Local Similarity 89.1%; Pred. No. 1,7e-23;
Matches 57; Conservative 1; Mismatches 1; Indels 5; Gaps 1;
OY 1 MRGSHHHHGS-----MASGDLKNKYAQLKRYKRSLKDKRAAEKQVSRLENEIEDLKA 55

Db 1 MGSHHHHHSGDDDKWASGDLKKNVQJLKRVRSLKDKAAELKQVSRLENEIEDLKA 60
QY 56 KIGD 59
61 KIGD 64

RESULT 2

US-08-956-307B-17
; Sequence 17, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956.307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-8906
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-956-307B-17

Query Match 64.6%; Score 243.5; DB 3; Length 76;

Best Local Similarity 75.0%; Pred. No. 1.3e-19;
Matches 48; Conservative 10; Mismatches 1; Indels 5; Gaps 1;

QY 1 MGSHHHHHSGS---MASGDLKKNVQJLKRVRSLKDKAAELKQVSRLENEIEDLKA 55
Db 1 MGSHHHHHSGDDDKWASGDLKKNVQJLKRVRSLKDKAAELKQVSRLENEIEDLKA 60

QY 56 KIGD 59
61 KIGD 64

RESULT 3

US-08-452-592B-7
; Sequence 7, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.

; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Wordperfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452.592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068.948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-4510
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-452-592B-7

Query Match 58.6%; Score 221; DB 1; Length 84;

Best Local Similarity 97.9%; Pred. No. 3.7e-17;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKKNVQJLKRVRSLKDKAAELKQVSRLENEIEDLKA KIGD LNN 62
Db 2 GDLKKNVQJLKRVRSLKDKAAELKQVSRLENEIEDLKA KIGD LKN 48

RESULT 4

US-08-452-592B-8
; Sequence 8, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Wordperfect Version 6.1
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,592B
FILING DATE: 25 May 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/068,948
FILING DATE: 25 May 1993
ATTORNEY/AGENT INFORMATION:
NAME: Donahue, Richard J.
NAME: Lamming, John H.
NAME: Ranucci, Vincent J.
REGISTRATION NUMBER: 22,062 (Donahue)
REGISTRATION NUMBER: 34,857 (Lamming)
REGISTRATION NUMBER: 29,579 (Ranucci)
REFERENCE/DOCKET NUMBER: NA-1096D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-233-4510
TELEFAX: 508-233-5167
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-452-592B-8

Query Match 56.5%; Score 213; DB 1; Length 84;
Best Local Similarity 93.6%; Pred. No. 2.7e-16;
Matches 44; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDLKKNVAVOLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
DB 2 GDLKKNVAVOLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 48

RESULT 5
US-08-452-592B-9
Sequence 9, Application US/08452592B
Patent No. 5712366
GENERAL INFORMATION:
APPLICANT: McGrath, Kevin P.
APPLICANT: Kaplan, David L.
TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
TITLE OF INVENTION: Self-Assembling Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: U.S. Army Soldier Systems Command
STREET: AMSCC-CC (Patent Counsel)
CITY: Natick
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01760-5035
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: DOS 6.22
SOFTWARE: Wordperfect Version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,592B
FILING DATE: 25 May 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/068,948
FILING DATE: 25 May 1993
ATTORNEY/AGENT INFORMATION:
NAME: Donahue, Richard J.
NAME: Lamming, John H.
NAME: Ranucci, Vincent J.
REGISTRATION NUMBER: 22,062 (Donahue)
REGISTRATION NUMBER: 34,857 (Lamming)
REGISTRATION NUMBER: 29,579 (Ranucci)
REFERENCE/DOCKET NUMBER: NA-1096D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-233-4510

TELEFAX: 508-233-5167
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-452-592B-9

Query Match 54.4%; Score 205; DB 1; Length 84;
Best Local Similarity 89.4%; Pred. No. 1.9e-15;
Matches 42; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDLKKNVAVOLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
DB 2 GDLKKNVAVOLKRRVSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 48

RESULT 6
US-08-956-307B-2
Sequence 2, Application US/08956307B
Patent No. 6090911
GENERAL INFORMATION:
APPLICANT: Petka, Wendy A.
APPLICANT: Titrrell, David A.
APPLICANT: Kevin P. McGrath
TITLE OF INVENTION: REVERSIBLE HYDROGELS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,307B
FILING DATE: 22-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07880/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-956-307B-2

Query Match 52.8%; Score 199; DB 3; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.8e-15;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 SGDLKKNVAVOLKRRVSLKDKAAELKQEVSRLENEIEDLKAK 56
DB 1 SGDLKKNVAVOLKRRVSLKDKAAELKQEVSRLENEIEDLKAK 42

RESULT 7


```

; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-956-307B-1

Query Match
Best Local Similarity 76.2%; Score 159; DB 3; Length 42;
Matches 32; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 15 SGD LKNKVAQLKRRVSLKDKAELKQVSRLENEIEDLKA 56
DB 1 SGD LENEVAQLEREVSRLEDEAAELKQVSRLENEIEDLKA 42

RESULT 10
US-08-452-592B-11
; Sequence 11, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; SELF-ASSEMBLING PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068,948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-4510
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-452-592B-11

Query Match
Best Local Similarity 72.9%; Score 124; DB 1; Length 42;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```

QY 16 GDLKKNVAQLKRRVSLKDKAELKQVSRLENEIEDLKA 55
DB 2 GDLKKNVAQLKRRVSLKDKAELKQVSRLENEIEDLKA 41

RESULT 11
US-08-956-307B-4
; Sequence 4, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-956-307B-4

Query Match
Best Local Similarity 24.1%; Score 91; DB 3; Length 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ISGLNNTSGIRPAKLN 74
DB 1 ISGLNNTSGIRPAKLN 18

RESULT 12
US-08-956-307B-3
; Sequence 3, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
```


OY 1 MRCSHNNHNN--GSMASGDILKKVQOLKRR---VRSLEDKRAELKOVSRLNEIEDLKA 55
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MGSSHHNNHHNSSLVGRSGHMDEALDPRRALDIYSANVDASHKEKELLIALDVLKLKLS 60

OY 56 K 56
|
Db 61 K 61

US-08-578-649-14
US-08-578-649-14
Sequence 14, Application US/08578649
Patent No. 5,770,366

GENERAL INFORMATION:
APPLICANT: Ulrich Bogdan
APPLICANT: Reinhard Butcher
TITLE OF INVENTION: BRILITTLE KALUA
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Palfie & Lynch
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 Inch 1.44 MB storage diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578,649
FILING DATE: 29-July-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 43 24 247.2
FILING DATE: 20-July-1993
ATTORNEY/AGENT INFORMATION:
NAME: Andrew L. Tligeloff
REGISTRATION NUMBER: 31,575
REFERENCE/DOCKET NUMBER: BOER 1035-PFF/ALT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-578-649-14

```

Query Match      21.8%   Score 82; DB 1; Length 18;
Best Local Similarity 76.5%   Pred. No. 0.005;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 MRGSHHHHHHSGSMASGD 17
    ||||| |::|::|::|::|
db 1 MKGSHHHHHHSGSYDDDD 17

```

Search completed: July 3, 2001, 14:56:47
Job time: 537 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:59:46 ; Search time 63.58 Seconds
(without alignments)
88,659 Million cell updates/sec

Title: US-09-490-291-11

Sequence: 377
1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRPRAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	25.1	1186	2	669708 chromosome segrega
2	89	23.6	315	3	JC7572 somite Maf1 protei
3	88	23.3	259	2	A70359 hydrogenase expres
4	86	22.8	155	2	I50167 B-G antigen - chic
5	85.5	22.7	323	2	I49529 transcription fact
6	85	22.5	321	2	D39371 Ig V-region-like B
7	85	22.5	880	2	F75103 conserved hypotet
8	85	22.5	979	2	J00894 P115 protein - MYC
9	84.5	22.4	281	2	F29316 hypotetrical prote
10	84	22.3	1199	2	T29145 tropomyosin I - sea
11	82	21.8	173	2	JC5610 hypotetrical prote
12	82	21.8	311	2	A56235 transcription acti
13	82	21.8	766	2	S37894 hypotetrical prote
14	82	21.8	1188	2	G83930 chromosome segrega
15	81.5	21.6	212	2	H69402 hypotetrical prote
16	81	21.5	166	2	S73342 hypotetrical prote
17	81	21.5	270	2	G71061 hypotetrical prote
18	81	21.5	396	2	T41405 hypotetrical prote
19	81	21.5	409	2	F72504 probable 26S prote
20	80.5	21.4	269	2	T18335 lcomg protein - Leg
21	80.5	21.4	647	2	A84265 hypotetrical prote
22	80	21.2	102	2	F64241 hypotetrical prote
23	80	21.2	103	2	B60608 myosin heavy chain
24	80	21.2	527	2	S33068 myosin heavy chain
25	80	21.2	1225	2	A43464 chromosome segrega
26	80	21.2	1940	2	A59287 myosin heavy chain
27	79.5	21.1	565	1	HMIVE3 hemagglutinin prec
28	79.5	21.1	565	1	HMIVE4 hemagglutinin prec
29	79	21.0	318	2	T49167 hypotetrical prote

30	79	21.0	369	1	TVFVAF transforming prote
31	79	21.0	407	2	F72343 hypotetrical prote
32	79	21.0	554	2	G72361 hypotetrical prote
33	78.5	20.8	564	1	HMIVE2 hemagglutinin prec
34	78.5	20.8	565	1	HMIVE2 hemagglutinin prec
35	78.5	20.8	1956	2	T16416 hypotetrical prote
36	78	20.7	456	2	E86903 hypotetrical prote
37	78	20.7	1169	2	A64505 P115 homolog - Met
38	78	20.7	1313	2	A48467 myosin heavy chain
39	78	20.7	1938	1	A40997 myosin heavy chain
40	77.5	20.6	348	2	T04618 heat shock protein
41	77.5	20.6	778	2	T30430 hypotetrical prote
42	77.5	20.6	1556	2	F96587 hypotetrical prote
43	77	20.4	292	2	I51171 transcription fact
44	77	20.4	338	2	I38587 retrovirus-related
45	77	20.4	1046	2	T42720 cytoplasmic linker

ALIGNMENTS

RESULT 1
669708
Chromosome segregation SMC protein - Bacillus subtilis
N/Alternate names: minichromosome stabilizing protein SMC
C/Species: Bacillus subtilis
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C/Accession: G69708; J04819; PC4029
R/Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allion, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Enlind, K.D.; Errington, J.; Farewell, C.; Ferrari,
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portee
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A/Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Dancho, A.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A/Reference number: A69580; M01D:98044033
A/Accession: G69708
A/Molecule type: DNA
A/Residues: 11186 <KUN>
A/Cross-references: GB:299112; GB:AL009126; NID:92633902; PIDN:CAB13467.1; PID:926339
A/Experimental source: strain 168
R/Oguro, A.; Kakeshita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.
Gene 172, 17-24, 1996
A/Title: The effect of Srb, a homologue of the mammalian SRP receptor alpha-subunit,
A/Reference number: J04819; M01D:96257247
A/Accession: J04819
A/Molecule type: DNA
A/Residues: 1-49, 'G', '51-161, 'G', '163-174, 'E', '176-177, 'G', '179-191, 'G', '193-227, 'P', '229-2
74, 'P', '476-493, 'D', '495-514, 'D', '516-541, 'V', '543-545, 'P', '547-585, 'SKPLRGSGPAFTISF', 601
, 'E', '741-1186 <OGU>
A/Cross-references: DDBJ:D64116; NID:91389548; PIDN:BAI10977.1; PID:91237015
R/Oguro, A.; Kakeshita, H.; Honda, K.; Takamatsu, H.; Nakamura, K.; Yamane, K.
DNA Res. 2, 95-100, 1995
A/Title: Srb, a Bacillus subtilis gene encoding a homologue of the alpha-subunit of t
A/Reference number: J04093; M01D:96093930
A/Accession: PC4029
A/Molecule type: DNA
A/Residues: 1171-1186 <OG2>
A/Cross-references: DDBJ:D49761; NID:91237017; PIDN:BAI08615.1; PID:91237018
C/Genetics:
A/Genes: smc
C/Superfamily: chromosome segregation protein SMC1

Query Match 25.1%, Score 94.5; DB 2; Length 1186;
Best Local Similarity 36.8%; Pred. No. 2.2;
Matches 25; Conservative 11; Mismatches 21; Indels 11; Gaps 2;
OY 11 GSMASGDLKNVAOLKRRVRSIKD-----KAELEKQVSRLENEIEDLKAKIGDL 60
DB 658 GSMGGAVKKKNNLSILGRSLELDVTKRLAEMEKTKALLEDEVTKLKHSIDOMKMLADL 717
OY 61 NNT-SGIR 67
DB 718 RETGEGLR 725
RESULT 2
JC7572
somite Maf1 protein - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
R:Kajihara, M.; Kawachi, S.; Kobayashi, M.; Ogino, H.; Takahashi, S.; Yasuda, K.
J. Biochem. 129, 139-146, 2001
A:Title: Isolation, characterization, and expression analysis of zebrafish large Mafs.
A:Reference number: JC7572; MUID: 21064923
A:Accession: JC7572
A:Molecule type: mRNA
A:Residues: 1-315 <KAU>
A:Cross-references: DDBJ:AB006324
C:Comment: This protein, a member of large Maf protein subfamily, containing a basic-leu
important roles in somitogenesis, and has specific functions as to the determination and
C:Genetics:
A:Gene: s-maf1
Query Match 23.6%; Score 89; DB 3; Length 315;
Best Local Similarity 25.5%; Pred. No. 1.7;
Matches 24; Conservative 12; Mismatches 20; Indels 38; Gaps 2;
OY 5 HHHHHGSMASGDL-----KNKVAOLKRRVRSIKDKAAE- 38
DB 179 HHHHHGHNHNAHRLDRPSDEQLVMTVRELNRQLRGFSKEEVITLKQKRRTLKRGYAAQ 238
OY 39 -----LKQEVSRLENEIEDLKAKIGDL 60
DB 239 SCRYKRVQORHMLESEKCTLOQVEQLKODVARL 272
RESULT 3
A70359
hydrogenase expression/formation protein B - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 24-Sep-1999
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70359; MUID: 98196666
A:Accession: A70359
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-259 <AAQ>
A:Cross-references: GB:AE000701, NID:g2983260; PIDN:MAC06860.1; PID:g2983266; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: hypB
C:Superfamily: hydrogenase expression/formation protein hypB
Query Match 23.3%; Score 88; DB 2; Length 259;
Best Local Similarity 28.6%; Pred. No. 1.7;
Matches 26; Conservative 12; Mismatches 29; Indels 24; Gaps 3;
OY 5 HHHHHGSMASGDLKNVAOLKRRVRSIKDKAAELKQ-----EVS 44

DB 19 HHHHHHTPALGD--KRVVEYLKTLISANDQAESNRHFERHGLAVNLMSPPSGKTT 76
OY 45 RLENEIEDLK--AKIGDLNNTSGIRPAKL 73
DB 77 LLERTIELLKDELKIGVTEGDLFTRDAERI 107
RESULT 4
I50167
B-g antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I50167
R:Kaufman, J.; Salomonsen, J.; Skjodet, K.
Immunogenetics 30, 440-451, 1989
A:Title: B-g cDNA clones have multiple small repeats and hybridize to both chicken MH
A:Reference number: I50161; MUID:9007532
A:Accession: I50167
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-155 <KAU>
A:Cross-references: GB:M27668; NID:g211253; PIDN:AAA48624.1; PID:g211254
Query Match 22.8%; Score 86; DB 2; Length 155;
Best Local Similarity 27.5%; Pred. No. 1.5;
Matches 22; Conservative 17; Mismatches 23; Indels 18; Gaps 2;
OY 7 HHHHSMASGDLKNVA-----QLKRVVSLDKAAELKQVSRLENEIED 52
DB 19 HCNEMDLISAADLKQIALVEQREGVEEMHITQLRKHVEKLSRAANLTKQKLENEIEE 78
OY 53 LKAKIGDLNNTSGIRPAKL 72
DB 79 VEKHL-----KITGRAPNLK 94
RESULT 5
I49529
transcription factor-kr - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 16-Jul-1999
C:Accession: I49529
R:Cordees, S.P.; Barsh, G.S.
Cell 79, 1025-1034, 1994
A:Title: The mouse segmentation gene kr encodes a novel basic domain-leucine zipper t
A:Reference number: A55200; MUID:95094266
A:Accession: I49529
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-323 <RES>
A:Cross-references: GB:I36435; NID:g625043; PIDN:AAA65689.1; PID:g625044
C:Superfamily: maf transforming protein; maf homology
C:Keywords: leucine zipper; transcription factor
F:212-301/Domain: maf homology <MAF>
Query Match 22.7%; Score 85.5; DB 2; Length 323;
Best Local Similarity 23.1%; Pred. No. 3.4;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;
OY 5 HHHHH-----GSMAS----- 15
DB 162 HHHHHHNASPPSSAASPAQAQLPTSHPGRPATATAAGNGSVEDRFSDQLVMSV 221
OY 16 -----GDLKNVAOLKRRVRSIKDK-----AAELK 40
DB 222 RELNHLRLGFTKDEVIYLRKQKRRTLKRGYAOSCRKYKVOQKHNLENEKTQILQOVEDLK 281
OY 41 QEVSRLENEIEDLKAKIGDLNNTSGIR 67
DB 282 DEVSRRLAREDAIKVKECEKLAN-SGFR 307

```
RESULT 6
D39371
Ig V-region-like B-G antigen 17 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 20-Mar-1992 #sequence_revision 10-Apr-1992 #text_change 21-Jul-2000
C:Accession: D39371
R:Miller, M.W.; Goto, R.; Young, S.; Chirivella, J.; Hawke, D.; Miyada, C.G.
Proc. Natl. Acad. Sci. U.S.A. 88, 4377-4381, 1991
A:Title: Immunoglobulin variable-region-like domains of diverse sequence within the major
A:Reference number: A39371; MIMID:91239571
A:Accession: D39371
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-321 <MIL>
A:Cross-references: GB:M61862; NID:g211259; PIDN:AAA48627.1; PID:g211261

Query Match 22.5%; Score 85; DB 2; Length 321;
Best Local Similarity 26.8%; Pred. No. 3.7;
Matches 22; Conservative 18; Mismatches 24; Indels 18; Gaps 2;

Oy 7 HHHGSASGDLK-----NKVAQLKRVKSLKDKAAELKQEVSRLENIED 52
Db 185 HLAERDSTADLKLAKLVEQREAVEERDSQLKQYKELGSATNLTQLKLENIIE 244
Oy 53 LKAKIGDLNMTSGIRPAKLN 74
Db 245 VEKHLAKI-----GIRAPNKLKH 262

RESULT 7
F75103
conserved hypothetical protein PAB0812 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Feb-2001
C:Accession: F75103
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: F75103
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KAM>
A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB50131.1; PID:g545866
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0812
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 22.5%; Score 85; DB 2; Length 880;
Best Local Similarity 37.2%; Pred. No. 10;
Matches 16; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Oy 18 LKNVAVQLKRVKSLKDKAAELKQEVSRLENIEDLKAKIGDL 60
Db 240 IKGKISLKIQVEKLRKGRKGLKLEKIVQIERSTIEKKAKISEL 282

RESULT 8
JQ0894
p115 protein - Mycoplasma hyorhinis
C:Species: Mycoplasma hyorhinis
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 02-Feb-2001
C:Accession: JQ0894
R:Notariicola, S.M.; McIntosh, M.A.; Wise, K.S.
Gene 97, 77-85, 1991
A:Title: A Mycoplasma hyorhinis protein with sequence similarities to nucleotide-binding
A:Reference number: JQ0894; MIMID:91138990
A:Accession: JQ0894
```

```
A:Molecule type: DNA
A:Residues: 1-979 <NOT>
A:Cross-references: GB:M34956
A:Note: the authors translated the codon AAA for residue 956 as Leu
C:Comment: This protein is located in the cytoplasm.
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: chromosome segregation protein SMCI
C:Keywords: nucleotide binding; P-loop
F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match 22.5%; Score 85; DB 2; Length 979;
Best Local Similarity 30.9%; Pred. No. 11;
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

Oy 17 DLKKNVAVQLKRVKSL-----KDKAAELKQEVSRLENIEDLKAKIGDLNMTSGI--- 66
Db 323 DQKTKIEIKQVSLKIQINASKQREIELDQUTRLNANKANSIKLDENDINKKIGVLE 382
Oy 67 RRPAPAKLN 74
Db 383 KKSAAAN 390

RESULT 9
F75216
hypothetical protein PAB2181 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: F75216
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: F75216
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-281 <KAM>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49181.1; PID:g545
C:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2181

Query Match 22.4%; Score 84.5; DB 2; Length 281;
Best Local Similarity 40.8%; Pred. No. 3.6;
Matches 20; Conservative 10; Mismatches 12; Indels 7; Gaps 1;

Oy 17 DLKKNVAVQLKRVKSL-----KVSRLDKAAELKQEVSRLENIEDLKAKIG 58
Db 222 ELERKVSLESLNEVETKVKSLKKEKELENKVKLEEEVKNLKEGIG 270

RESULT 10
T29145
hypothetical protein F56A3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29145
R:Geisel, C.; Bradshaw, H.; Wamsley, P.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid F56A3.
A:Reference number: 220578
A:Accession: T29145
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1199 <GEI>
A:Cross-references: EMBL:U60446; PIDN:AAB37802.1; GSPDB:GN00019; CESP:F56A3.4
A:Experimental source: strain Bristol N2; clone F56A3
C:Genetics:
A:Gene: CESP:F56A3.4
A:Map position: 1
```


Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20263314
A:Accession: G83960
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 11188 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; MID:g10174886; PIDN:BA06206.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: smc

Query Match 21.8%; Score 82; DB 2; Length 1188;
Best Local Similarity 35.0%; Pred. No. 24;
Matches 21; Conservative 12; Mismatches 17; Indels 10; Gaps 1;

OY 11 GSMASGDLKHK-----VQLKRRVSLKDAELKQVSRLENEIEDLKAKIGDL 60
DB 659 GSMGGSVKOKPPLGRORLELTOKLEKAAKLEQEVKQKNERTELQQLIEDL 718

RESULT 15

H69402
hypothetical protein AF1225 - *Archaeoglobus fulgidus*
C:Species: *Archaeoglobus fulgidus*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: H69402
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artlisch, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: H69402
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <KLE>
A:Cross-references: GB:AE001019; GB:AE000782; MID:g2689342; PIDN:AAB90025.1; PID:g264936

Query Match 21.6%; Score 81.5; DB 2; Length 212;
Best Local Similarity 31.1%; Pred. No. 4.7;
Matches 19; Conservative 15; Mismatches 20; Indels 7; Gaps 1;

OY 17 DLKRVKQQL-----KRVKSLKDAELKQVSRLENEIEDLKAKIGDUNNTSGIRRP 69
DB 130 DLONKARMLGIONELQKQIEELQKLEKRAKTELQKRRKSDLELEKIRLEINROSSIYRL 189

OY 70 A 70
DB 190 A 190

Search completed: July 3, 2001, 14:59:47
Job time: 372 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:51 ; Search time 37.35 Seconds
(without alignments)
67.869 Million cell updates/sec

Title: US-09-490-291-11

Sequence: 1 MRGSHHHHSGMASGDLKN.....AKIGDLNNTSGTRPAAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	94.5	25.1	1186	1 SMC_BACSU	P11834 bacillus su
2	86.5	22.9	323	1 SMC_BACSU	P54842 bacterium norv
3	86	22.8	548	1 CEAK_ECOLI	Q47502 escherichia
4	85.5	22.7	323	1 MAR1_MOUSE	P34841 mus musculus
5	85	22.5	979	1 P115_MYCHR	P15108 mycoplasma
6	82	21.8	766	1 STB6_YEAST	P16085 saccaromyc
7	81	21.5	166	1 YB38_MYCPN	P5260 mycoplasma
8	80	21.2	102	1 Y334_MERJA	O5780 methanococ
9	80	21.2	1225	1 SMC1_YEAST	P32908 saccaromyc
10	79.5	21.1	365	1 HEMA_IHAHL	P16994 saccaromyc
11	79.5	21.1	565	1 HEMA_IHAHL	P17000 influenza a
12	79	21.0	369	1 TMAF_AVISA	P33091 avian muscu
13	78.5	20.8	564	1 HEMA_IAGRE	P19698 influenza a
14	78.5	20.8	565	1 HEMA_IAGRE	P15658 influenza a
15	78	20.7	1938	1 MYS_AEOTR	P24723 aequipecten
16	78	20.7	4349	1 DYHC_FUSSO	P78716 fusarium so
17	77	20.4	461	1 US45_LACLC	P22865 lactococcus
18	76.5	20.3	741	1 BSG2_DROME	P11929 drosophila
19	75.5	20.0	550	1 HEMA_IADHK	P13257 influenza a
20	75.5	20.0	1509	1 MYSN_ACACN	P05659 acanthamoeb
21	75	19.9	715	1 CLPB_MYCPN	P75247 mycoplasma
22	74.5	19.8	830	1 PAM1_YEAST	P7304 saccaromyc
23	74.5	19.6	879	1 MYSF_DROME	P35415 drosophila
24	74	19.6	539	1 MYS3_HYDAT	P39922 hydra atten
25	74	19.6	1957	1 YD86_SCHPO	Q10411 schistosom
26	74	19.6	2022	1 ANTI_ONCVO	P11249 onchocerca
27	73.5	19.5	550	1 HEMA_IABAN	P03441 influenza a
28	73.5	19.5	550	1 HEMA_IADH1	P12532 influenza a
29	73.5	19.5	550	1 HEMA_IADH2	P12533 influenza a
30	73.5	19.5	550	1 HEMA_IADH3	P12534 influenza a
31	73.5	19.5	550	1 HEMA_IADH4	P12535 influenza a
32	73.5	19.5	550	1 HEMA_IADH6	P12537 influenza a
33	73.5	19.5	550	1 HEMA_IADH7	P12538 influenza a

ALIGNMENTS

RESULT	1	STANDARD	PRT	1186 AA.
SMC_BACSU				
ID	SMC_BACSU			
AC	P51834; 031735;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	CHROMOSOME PARTITION PROTEIN SMC.			
GN	SMC.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RA	Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;			
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1171-1186 FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=96093930; PubMed=7584053;			
RA	Oguro A., Kakeshita H., Honda K., Takamatsu H., Nakamura K.,			
RA	Yamane K.;			
RT	"srb: a Bacillus subtilis gene encoding a homologue of the alpha-			
RT	subunit of the mammalian signal recognition particle receptor.";			
RT	DNA Res. 2:95-100(1995).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=98367134; PubMed=9701812;			
RA	Moriya S., Tsujikawa E., Hassan A.K., Asai K., Kodama T.,			
RA	Ogasawara N.;			
RT	"A Bacillus subtilis gene-encoding protein homologous to eukaryotic			
RT	SMC motor protein is necessary for chromosome partition.";			
RT	Mol. Microbiol. 29:179-187(1998).			
RN	[5]			
RP	CHARACTERIZATION.			
RX	MEDLINE=98241483; PubMed=9573042;			
RA	Britton R.A., Lin D.C., Grossman A.D.;			
RT	"Characterization of a prokaryotic SMC protein involved in chromosome			
RT	partitioning.";			
RT	Genes Dev. 12:1254-1259(1998).			
CC	-1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND			
CC	PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION.			
CC	-1- SIMILARITY: BELONGS TO THE SMC FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D64116; BAA10977.1; -

DR EMBL; 299112; CAB13467.1; -

DR EMBL; D49781; BAA08615.1; -

DR Subtilist; Bg11538; smc.

KW ATP-binding; Coiled coil.

FT NP_BIND 31 38

FT DOMAIN 244 481

FT 664 943

FT 989 1031

FT 50 50

FT 162 162

FT 175 175

FT 178 178

FT 192 192

FT 228 228

FT 236 236

FT 264 264

FT 271 271

FT 284 284

FT 309 309

FT 342 342

FT 365 365

FT 438 438

FT 444 444

FT 475 475

FT 494 494

FT 515 515

FT 542 542

FT 546 546

FT 586 600

FT CONFLICT 623 632

FT CONFLICT 664 664

FT CONFLICT 676 676

FT CONFLICT 680 680

FT CONFLICT 694 694

FT CONFLICT 701 701

FT CONFLICT 726 726

FT CONFLICT 738 740

FT CONFLICT 1186 AA; 135510 MW; 0163227AATFBCA2B CRC64;

Query Match

Best Local Similarity 25.1%; Score 94.5; DB 1; Length 1186;

Matches 25; Conservative 11; Mismatches 21; Indels 11; Gaps 2;

QY 11 GSMASDLKKNVAQLKRVSLKDK-----KAELEKQEVSRLENEIEDIKAKIGDL 60

DB 658 GSMTGAVKVKNNKSLGRSELEDVTKRLAEWEKTKALLDEQVKTKHSIQMEKKLADL 717

QY 61 NNT-SGIR 67

DB 718 RETGEGLR 725

RESULT 2

MAP1_RAT STANDARD; PRT; 323 AA.

AC P54842;

DT 01-OCT-1996 (Rel. 34, Created)

DR 01-OCT-1996 (Rel. 34, Last sequence update)

DE 01-OCT-2000 (Rel. 40, Last annotation update)

DE TRANSCRIPTION FACTOR MAP1.

GN MAPB OR MAP1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxId=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-WISTAR; TISSUE-Liver;

RA MEDLINE=97190228; PubMed=9038383;

RA Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,

RA Kuboki Y., Nishizawa M., Nishi S.;

RT "Rat mat related genes: specific expression in chondrocytes, lens and

RT spinal cord.";

RL Oncogene 14:745-750(1997).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- SIMILARITY: BELONGS TO THE B2IP FAMILY. MAP SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U56241; AAB50062.1; -

DR HSSP; P05412; LUN.

KW Transcription regulation; DNA-binding; Nuclear protein.

FT DOMAIN 131 143

FT 158 167

FT 194 201

FT 238 264

FT 266 287

FT 323 AA; 35792 MW; 6E386340D1F840A5 CRC64;

SO SEQUENCE

Query Match

Best Local Similarity 22.9%; Score 86.5; DB 1; Length 323;

Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;

QY 5 HHHHH-----GSMAS----- 15

DB 162 HHHHHQASPPSSAASPAQQLPTSHPGPHAAATAAGSGVSDREFSDQLVMSV 221

QY 16 -----GDLKKNVAQLKRVSLKDK-----AAELK 40

DB 222 RELNHLRGFTDEVTIRLKKRRRTLKNQYASCRKRVQOKHLENEKTKLQGVQLK 281

QY 41 QEVSRLENEIEDIKAKIGDLNNTSGIR 67

DB 282 QEVSRLENEIEDIKAKIGDLNNTSGIR 307

RESULT 3

CEAK_ECOLI STANDARD; PRT; 548 AA.

AC Q47502; P75615;

DT 15-DEC-1998 (Rel. 37, Created)

DR 15-DEC-1998 (Rel. 37, Last sequence update)

DE 01-OCT-2000 (Rel. 40, Last annotation update)

DE COLICIN K.

GN CKA.

OS Escherichia coli.

OC Plasmid ColK-K235.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxId=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K49;

RX MEDLINE=96074330; PubMed=7592493;

RA Pilai H., Braun V.;

RT "Strong function-related homology between the pore-forming colicins K

RT and S.";

RT J. Bacteriol. 177:6973-6977(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA Izard J., Chartier M., Baty D.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: THIS COLICIN IS A CHANNEL-FORMING COLICIN. THIS CLASS OF

```

CC TRANSMEMBRANE TOXINS DEPOLARIZE THE CYTOPLASMIC MEMBRANE, LEADING
CC TO DISSIPATION OF CELLULAR ENERGY.
CC -1- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
CC -1- SIMILARITY: BELONGS TO THE CHANNEL FORMING COLICIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X87834; CA61099.1; -.
CC EMBL: U27452; AAB41288.1; -.
CC InterPro: IPR000293; -.
CC Pfam: PF01024; Colicin; 1.
CC PRINTS: PR00280; CHANCOLICIN.
CC PROSITE: PS00276; CHANNELCOLICIN; 1.
CC Antidiotic: Bacteriocin; Plasmid; Transmembrane.
CC TRANSMEM 505 525 POTENTIAL.
CC FT CONFLICT 155 155 S -> G (IN REF. 2).
CC SEQUENCE 548 AA; 59661 MW; 2E67689D088CCE86 CRC64;
SO
Query Match 22.8%; Score 86; DB 1; Length 548;
Best Local Similarity 32.1%; Pred. No. 1.8;
Matches 18; Conservative 13; Mismatches 15; Indels 10; Gaps 1;
OY 6 HHHHGSMAAGDLKKNVAOLKRRVRSLSKDAELKOEVSRLNTEEDLAKIGDGLN 61
DB 188 HHSYHG-----ITDLERVDYDLQKRSNLDADISLNSYKNTLQSKIGDYN 233
RESULT 4
MAPL_MOUSE STANDARD; PRT; 323 AA.
AC P54841;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTION FACTOR MAP1 (SEGMENTATION PROTEIN KR) (KREISLER).
GN MAPB OR MAP1 OR KRAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95094266; Pubmed=8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor.";
RL Cell 79:1025-1034(1994).
CC -1- FUNCTION: MAY PLAY AN EARLY ROLE IN AXIAL PATTERNING (HINDBRAIN
CC SEGMENTATION).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, GUT, LUNG, AND BRAIN.
CC -1- DEVELOPMENTAL STAGE: DETECTABLE AT 8.0 DPC (ONE SOMITE) AS A BAND
CC IN THE CAUDAL HINDBRAIN, AND BY 8.5 DPC (SIX TO EIGHT SOMITES),
CC THE HIGH LEVEL DOMAIN EXHIBITS A SHARP ROSTRAL EDGE COINCIDENT
CC WITH THE R4/R5 BOUNDARY AND A DIFFUSE CAUDAL EDGE LOCATED MIDWAY
CC THROUGH R6.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: L36435; AAA65689.1; -.
CC DR HSSP; P05412; IJUN.
CC DR TRANSFAC; T01439; -.
CC DR MGD; MGI:104555; MafB.
CC KW Transcription regulation; DNA-binding; Nuclear protein.
CC FT DOMAIN 131 143
CC FT DOMAIN 158 167 POLY-HIS.
CC FT DNA_BIND 238 264 BASIC MOTIF.
CC FT DOMAIN 266 287 LEUCINE-ZIPPER.
CC FT NOTAGEN 248 248 N-S; LOSS OF TRANSCRIPTIONAL ACTIVITY.
CC SEQUENCE 323 AA; 35809 MW; D77AE07ABD9C2AD2 CRC64;
SO
Query Match 22.7%; Score 85.5; DB 1; Length 323;
Best Local Similarity 23.1%; Pred. No. 1.1;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;
OY 5 HHHHH-----GSMAS-----
DB 162 HHHHHQASPPSSAASPAQQLPTSHQPGPHATTAAGNGSVEDRSDDQVMSV 221
OY 16 -----GDLKNRYAOLKRRVRSLSKDK-----AAELK 40
DB 222 RELNRHLRGFTKDEVYRLKQKRTLNKRGYAGSCRYKRVQKHHLENEKQLIQVQDLK 281
OY 41 QEVSRLNEIEDLAKIGDGLNNTSGIR 67
DB 282 QEVSRLARENDAYKVEKELAN-SGFR 307
RESULT 5
P115_MYCHR STANDARD; PRT; 979 AA.
AC P41508;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE P115 PROTEIN.
OS Mycoplasma hyorhinis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91138990; Pubmed=1825306;
RA Notariicola S.M., McIntosh M.A., Wise K.S.;
RT "A Mycoplasma hyorhinis protein with sequence similarities to
RT nucleotide-binding enzymes.";
RL Gene 97:77-85(1991).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SIMILAR TO OTHER MYCOPLASMA
CC P115.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M34956; AAA25423.1; -.
CC DR PIR; JQ0894; J00894.
CC KW ATP-binding; Coiled coil.
CC FT NP_BIND 32 39
CC FT DOMAIN 169 224 COILED COIL (POTENTIAL).
CC FT DOMAIN 231 400 COILED COIL (POTENTIAL).
CC FT DOMAIN 569 821 COILED COIL (POTENTIAL).
CC FT DOMAIN 884 912 ALA/ASP-RICH (DA-BOX).
CC SEQUENCE 979 AA; 110566 MW; 30D51C56B56280F4 CRC64;

```

```

Query Match          22.5%; Score 85; DB 1; Length 979;
Best Local Similarity 30.9%; Pred. No. 4;
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

QY 17 LKKNVAQLKRRVRSI-----KDKAELKQEVSRLENEIEDLAKIDLNNTSGI----66
DB 323 DQKTKIEEIKQVESIKIQINASKOREIELDQDLTRLNKANSIKLQENDINKEIGVLE 382

QY 67 RRPAAKLN 74
DB 383 KKSAAAN 390

RESULT 6
STB6_YEAST
ID STB6_YEAST STANDARD; PRT; 766 AA.
AC P36085;
DT 01-JUN-1994 (Rel. 29, Created)
DR 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE STB6 PROTEIN
GN STB6 OR YKL072W OR YKL352.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins."
RL yeast 10:569-574(1994).
RN [2]
RP SEQUENCE OF 1-557 FROM N.A.
RA Pohl T.M., Pohl F.M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98055153; PubMed=9393435;
RA Kaestlen M.M., Stillman D.J.;
RT "Identification of the Saccharomyces cerevisiae genes STB1-STB5
RT encoding Sin3p binding proteins."
RL Mol. Gen. Genet. 256:376-386(1997).
CC -1- FUNCTION: BINDS TO SIN3.
CC -1- SIMILARITY: STRONG, TO YEAST STB2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC
CC EMBL; X75780; CAA53402.1; -
DR EMBL; 228072; CAA81909.1; -
DR PIR; S37894; S37894.
DR PIR; S39170; S39170.
DR PIR; S44515; S44515.
DR SGD; S000155; YKL072W.
SQ SEQUENCE 766 AA; 88835 MW; 9BCBEA2EE03A9AF1 CRC64;

Query Match          21.8%; Score 82; DB 1; Length 766;
Best Local Similarity 26.7%; Pred. No. 5;
Matches 20; Conservative 20; Mismatches 25; Indels 10; Gaps 2;

```

```

QY 2 RGSHHH-HHGSMA-----GDLKNVAQLKRRVRSIKDKAELKQEVSRLENEIE 51
DB 620 RGSNEHWEYGNMNOQLIESEKKDKMSLKKVNYILDRVADVBSVDHFDKRLK 679

QY 52 DLKAKIGDLNNTSGI 66
DB 680 DVRKILLEQNNNSKDI 694

RESULT 7
YB38_MYCPN
ID YB38_MYCPN STANDARD; PRT; 166 AA.
AC P75260;
DT 01-OCT-2000 (Rel. 40, Created)
DR 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MPN138 (E07_ORF166).
GN MPN138 OR MP016.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE UPF0134 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC
CC EMBL; AE000003; AAB95664.1; -
DR InterPro; IPR002862; -
DR Pfam; PF01519; DUF16; 1.
KW Hypothetical protein.
SQ SEQUENCE 166 AA; 19518 MW; BE44F5377B2FA709 CRC64;

Query Match          21.5%; Score 81; DB 1; Length 166;
Best Local Similarity 37.2%; Pred. No. 1;
Matches 16; Conservative 15; Mismatches 12; Indels 0; Gaps 0;

QY 18 LKKNVAQLKRRVRSIKDKAELKQEVSRLENEIEDLAKITGL 60
DB 81 IENKVDKLEVKVDKLEKVKLEAKVDKLEKVDKLEAKVDKL 123

RESULT 8
Y334_METJA
ID Y334_METJA STANDARD; PRT; 102 AA.
AC Q57780;
DT 01-NOV-1997 (Rel. 35, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0334.
GN MJ0334.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE-96337999; PubMed-8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merriam J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Puhmann J.L., Nguyen D.,
 RA Uiterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*.";
 RL Science 273:1058-1073(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 CC EMBL: U67487; AAB98322.1; -
 CC TIGR: M0334; -
 CC Hypothetical protein.
 CC
 CC DOMAIN 60 89 GLU/LYS-RICH.
 CC SEQUENCE 102 AA; 11555 MW; 1C9A45826A0B2AE3 CRC64;
 FT
 SQ
 Query Match 21.2%; Score 80; DB 1; Length 102;
 Best Local Similarity 40.0%; Pred. No. 0.96; Mismatches 17; Indels 8; Gaps 1;
 Matches 22; Conservative 8; Mismatches 17; Indels 8; Gaps 1;
 QY 19 KKKVAQLKKRVK-----SLKDKAELKQVSRLENEIEDLAKIGDLNNTSG 65
 Db 37 KDKVDALVRKHEMOPFHEVSGNKKIAELKAKELAKKEKEIELEIKETIGLKGKG 91
 ID SMCI_YEAST STANDARD; PRT; 1225 AA.
 AC P32908;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CHROMOSOME SEGREGATION PROTEIN SMCI (DA-BOX PROTEIN SMCI).
 GN SMCI OR CHL10 OR YFL008W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 ON NCBI_TaxID=4932;
 OX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94103320; PubMed-8276886;
 RA Strunikhov A.V., Laktionov V.L., Koshland D.;
 RT "SMCI, an essential yeast gene encoding a putative head-rod-tail
 RT protein is required for nuclear division and defines a new ubiquitous
 RT protein family.";
 RL J. Cell Biol. 123:1635-1648(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE-95400292; PubMed-7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT *Saccharomyces cerevisiae*.";
 RL Nat. Genet. 10:261-268(1995).
 CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
 CC PART OF A CHROMOSOME CONDENSATION MOTOR.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.

CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
 CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 CC EMBL: L00602; AAA16595.1; -
 CC DR EMBL: D50617; BAA09230.1; -
 CC DR EMBL: D31600; BAA06496.1; -
 CC DR PIR: A49464; A49464.
 CC DR PIR: S41804; S41804.
 CC DR SCD: S0001886; SMCI.
 CC KM Mitosis; ATP-binding; Coiled coil; Nuclear protein.
 CC NP_BIND 33 40
 CC DOMAIN 173 489
 CC DOMAIN 679 1063 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1057 1061 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1137 1164 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC FT MUTAGEN 173 173 ALA/ASP-RICH (DA-BOX).
 CC FT MUTAGEN 458 458 S->L: IN TS MUTANT SMCI-2.
 CC FT MUTAGEN 458 458 N->D: IN TS MUTANT SMCI-1.
 CC SEQUENCE 1225 AA; 141279 MW; B504017AA0ECCA8C CRC64;
 FT
 SQ
 Query Match 21.2%; Score 80; DB 1; Length 1225;
 Best Local Similarity 33.9%; Pred. No. 13; Mismatches 21; Indels 4; Gaps 1;
 Matches 19; Conservative 12; Mismatches 21; Indels 4; Gaps 1;
 QY 12 SMASGDLKKVAQLKKRVSLKDKAELKQVSRLENEIEDLAKIGDLNNT 63
 Db 718 SLMSDIALNLTQYTGQKRSLENNRLKYNHNDLEIKETIGLKKKKDLDEMT 773
 ID HEMA_IHAHL STANDARD; PRT; 565 AA.
 AC P16994; O83991; O83990;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HAI CHAIN;
 DE HEMAGGLUTININ HAI CHAIN].
 GN HA.
 OS Influenza A virus (strain A/Egypt/Algiers/772).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;
 OC Influenza A virus.
 ON NCBI_TaxID=11393;
 OX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89204899; PubMed-2705299;
 RA Kawasaka Y., Bean W.J., Webster R.G.;
 RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
 RL Virology 169:283-292(1989).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
 CC (HAI AND HAI2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M24721; AAA43100.1; ALT_SEQ.
DR PIR; D34064; HMTVE4.
DR HSSP; P03437; IHMT.
DR InterPro; IPRO00149; -
DR InterPro; IPRO000386; -
DR InterPro; IPRO01364; -
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR PRINTS; PR00331; HEMAGGLUTN2.
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
KW SIGNAL
FT CHAIN 1 16
FT CHAIN 17 340 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 345 565 HEMAGGLUTININ HAZ CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63831 MW; BA533050DC3F186B CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;
Best Local Similarity 28.1%; Pred. No. 6.5;
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

OY 3 GSHHHHHHSMASGDLKKNVA-----OLKRRVSLKDKAELKOEVRLENEIEDLKA 55
DB 367 GRRHNSGEGTGAAGDLKSTQAALIDQINGKLNRYIEKTKNEKHOIEKSEFVGRIDDEK 426
OY 56 KIGD 59
DB 427 YVED 430

RESULT 11
HEMA_IAMTO STANDARD; PRT; 565 AA.
ID HEMA_IAMTO PRT; 565 AA.
AC P17000; 084002; 084003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HAI CHAIN;
HEMAGGLUTININ HAZ CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Tokyo/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OY NCBI_TaxID=11418;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8920489; PubMed=2705299;
RA Kawoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RT Virology 169:283-292(1989)
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24720; AAA43111.1; ALT_SEQ.

DR PIR; C34064; HMTVE3.
DR HSSP; P03437; IHMT.
DR InterPro; IPRO00149; -
DR InterPro; IPRO000386; -
DR InterPro; IPRO01364; -
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR PRINTS; PR00331; HEMAGGLUTN2.
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
KW SIGNAL
FT CHAIN 1 16
FT CHAIN 17 343 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 345 565 HEMAGGLUTININ HAZ CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63580 MW; 84B7D4AD70629B7A CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;
Best Local Similarity 28.1%; Pred. No. 6.5;
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

OY 3 GSHHHHHHSMASGDLKKNVA-----OLKRRVSLKDKAELKOEVRLENEIEDLKA 55
DB 367 GRRHNSGEGTGAAGDLKSTQAALIDQINGKLNRYIEKTKNEKHOIEKSEFVGRIDDEK 426
OY 56 KIGD 59
DB 427 YVED 430

RESULT 12
TMAF_AVIS4 STANDARD; PRT; 369 AA.
ID TMAF_AVIS4 PRT; 369 AA.
AC P23091;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSFORMING PROTEIN TMAF.
GN V-TMAF.
OS Avian musculoponeurotic fibrosarcoma virus AS42.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OC NCBI_TaxID=11873;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9004665; PubMed=2554284;
RA Nishizawa M., Kataoka K., Goto N., Fujiwara K.T., Kawai S.;
RT "v-maf, a viral oncogene that encodes a 'leucine zipper' motif.";
RT Proc. Natl. Acad. Sci. U.S.A. 86:7711-7715(1989).
CC -1- FUNCTION: MIGHT BE A TRANSCRIPTIONAL TRANS-ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: INDUCES MUSCULOAPONEUROTIC FIBROSARCOMA IN CHICKENS.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A ENV-TMAF
POLYPROTEIN.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M26769; AAA42377.1; -
DR PIR; B33975; TVEVAF.
DR HSSP; P05412; IJUN.

```

DR TRANSFAC: T01430;
KM Transferring protein; Transcription regulation; DNA-binding;
KW Nuclear protein.
FT DOMAIN 163 167 POLY-ALA.
FT DOMAIN 173 182 HIS-RICH.
FT DOMAIN 184 191 POLY-GLY.
FT DOMAIN 214 223 POLY-GLY.
FT DOMAIN 233 240 POLY-GLY.
FT DNA_BIND 274 300 BASIC MOTIF.
FT DOMAIN 302 323 LEUCINE-ZIPPER.
SQ SEQUENCE 369 AA: 38892 MW: F386B220ACE50FF6 CRC64;

Query Match 21.0%; Score 79; DB 1; Length 369;
Best Local Similarity 24.1%; Pred. No. 4.6;
Matches 26; Conservative 8; Mismatches 20; Indels 54; Gaps 2;

OY 3 GSHHHHSGMASGDL-----KNKYAOLKRYVSLDK 35
DB 226 GLNPHHGCGGGGGLHFDRESDQLVTMSRELNLQGLGVSKKEYIRLKKRRTLKNR 285
OY 36 -----AELKQVSRLENEIEDLKK 56
DB 286 GYAQSCFRKVRQHRVLESEKNQLDOVEHLKQEISSLVRERDAYKREK 333

RESULT 13
HEMA_IAGRE
ID HEMA_IAGRE STANDARD; PRT; 564 AA.
AC P19698;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HAI CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Grey teal/Australia/2/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBL_Taxid=11385;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Daniels R.O., Bean W.J., Kawaka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RT different regions of the world.";
RL Virology 169:408-417(1989).
RN -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
RN CELL RECEPTORS AND FOR INITIATING INFECTION.
RN -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
RN (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
RN -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M25284; AAA43217.1; -.
CC PIR: B34214; HMIVE2.
CC HSSP: P03437; 2HMG.
CC InterPro: IPR000149; -.
CC InterPro: IPR000386; -.
CC InterPro: IPR001364; -.
CC Pfam: PF00509; Hemagglutinin; 1.
CC PRINTS: PR00329; HEMAGGLUTIN1.
CC PRINTS: PR00330; HEMAGGLUTIN1.
CC PRINTS: PR00331; HEMAGGLUTIN2.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

```

```

FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 564 AA: 63237 MW: A4CD347146D0971C CRC64;

Query Match 20.8%; Score 78.5; DB 1; Length 564;
Best Local Similarity 28.1%; Pred. No. 7.9;
Matches 18; Conservative 14; Mismatches 25; Indels 7; Gaps 1;

OY 3 GSHHHHSGMASGDLKNKYA-----QLKRVRSLSKDKAELKQVSRLENEIEDLKA 55
DB 366 GFRHQNAEGTGTAADLSTQQAIDQINGKLRLIEKINDKHQIEKEFEVGEGRIDLKK 425
OY 56 KIGD 59
DB 426 YVED 429

RESULT 14
HEMA_IAHMI
ID HEMA_IAHMI STANDARD; PRT; 565 AA.
AC P15658; O83988; O83987; O67097; O67098;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HAI CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Miami/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBL_Taxid=11407;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
RN [2]
RN -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
RN CELL RECEPTORS AND FOR INITIATING INFECTION.
RN -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
RN (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
RN -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M24719; AAA43105.1; ALT_SEQ.
CC EMBL: M29257; AAA43164.1; -.
CC PIR: B34064; HMIVE2.
CC HSSP: P03437; 1HTM.
CC InterPro: IPR000149; -.

```


Fri Jul 6 08:02:47 2001

us-09-490-291-11.rsp

Page 9

This Page Blank (uspto)


```
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE TRANSCRIPTION FACTOR VAL.
GN VALENTINO.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165393; PubMed=9425134;
RA Moens C.B., Cordes S.P., Giorjani M.W., Barsh G.S., Kimmel C.B.;
RT "Equivalence in the genetic control of hindbrain segmentation in fish
RT and mouse.";
RL Development 125:381-391(1998).
DR EMBL: AF006641; AAC18821.1; -.
DR ZFIN: ZDB-GENE-980526-515; val.
DR InterPro: IPR001871; -.
DR SMART: SM00338; BRLZ; 1.
SQ SEQUENCE 356 AA; 40243 MW; 07420DB0F6CD08F1 CRC64;

Query Match 23.7%; Score 89.5; DB 13; Length 356;
Best Local Similarity 22.2%; Pred. No. 3.1;
Matches 28; Conservative 13; Mismatches 22; Indels 63; Gaps 3;

OY 5 HHHHHH-----GSM-----ASGDLKKVAOLKKR 28
    |||||
DB 215 HHHHHHHPGQGHGNGVGLNVEDRSDQLYMSVRELNRHLRGTKDQVIRLKK 274
    |||||
OY 29 VSLDK-----AAELKQVSRLENEDEDKAKIGDIN 61
    |||||
DB 275 RRTLNRYGAOSCRFRKRYQOKHLENEKTOLINOVOLKQEIINRLAREHDVAKLCEKIT 334
    |||||
OY 62 NTSGR 67
DB 335 GANGFR 340

RESULT 3
OY 091891 PRELIMINARY; PRT: 146 AA.
AC 091891:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE BZIP TRANSCRIPTION FACTOR L-MAF (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRED FAYOMI, AND BREED C-B1 LEHORN; TISSUE-SPLEEN;
RA Zhou H., Lamont S.J.;
RT "Genetic variation among chicken lines and mammalian species in
RT specific genes.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF221558; AAF71240.1; -.
DR EMBL: AF221557; AAF71239.1; -.
DR InterPro: IPR001871; -.
DR SMART: SM00338; BRLZ; 1.
FT NON_TER 1
FT 146
SQ SEQUENCE 146 AA; 17579 MW; 382BAD1D098E1DCC CRC64;

Query Match 23.5%; Score 88.5; DB 13; Length 146;
Best Local Similarity 27.5%; Pred. No. 1.6;
Matches 28; Conservative 11; Mismatches 14; Indels 49; Gaps 3;

OY 4 SHHHHHG-----SMASGDL-----KKVAOLKKRVSLDK----- 35
```

```
DB 39 AHHHHHHHLLREFRSDQVMSVRELNRQLRGFSKEEVLRLKQRRTLKNGVAOSC 98
    :|||||
OY 36 -----AAELKQVSRLENEDEDKAK 56
    :|||||
DB 99 RYKRVQQRHILENEKCOLQSOVEOLKQVSRLLAKERDLYKEK 140

RESULT 4
OY 057342 PRELIMINARY; PRT: 286 AA.
AC 057342:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE BZIP TRANSCRIPTION FACTOR MAFA.
GN MAFA.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RA Benkheila S., Felder-Schmittbuhl M.P., Calothy G.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF034693; AAC60377.1; -.
DR InterPro: IPR001871; -.
DR SMART: SM00338; BRLZ; 1.
SQ SEQUENCE 286 AA; 32463 MW; 0D4ED7BF08D91397 CRC64;

Query Match 23.5%; Score 88.5; DB 13; Length 286;
Best Local Similarity 27.5%; Pred. No. 3;
Matches 28; Conservative 11; Mismatches 14; Indels 49; Gaps 3;

OY 4 SHHHHHG-----SMASGDL-----KKVAOLKKRVSLDK----- 35
    :|||||
DB 157 AHHHHHHHLLREFRSDQVMSVRELNRQLRGFSKEEVLRLKQRRTLKNGVAOSC 216
    :|||||
OY 36 -----AAELKQVSRLENEDEDKAK 56
    :|||||
DB 217 RYKRVQQRHILENEKCOLQSOVEOLKQVSRLLAKERDLYKEK 258

RESULT 5
OY 066901 PRELIMINARY; PRT: 259 AA.
AC 066901:
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE HYDROGENASE EXPRESSION/FORMATION PROTEIN B.
GN HYB.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Nijay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
RL EMBL: AE000701; AAC06860.1; -.
DR InterPro: IPR000822; -.
DR InterPro: IPR002894; -.
DR Pfam: PF01495; Hyb_Ureg; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; UNKNOWN.1.
SQ SEQUENCE 259 AA; 29015 MW; 4522AC8CF53806C CRC64;
```

Query Match	Similarity	28.6%	Pred. No. 3:	23.3%	Score 88:	DB 2:	Length 259:
Best Local	Similarity	28.6% <td>Pred. No. 3:</td> <td>23.3% <td>Score 88:</td> <td>DB 2:</td> <td>Length 259:</td> </td>	Pred. No. 3:	23.3% <td>Score 88:</td> <td>DB 2:</td> <td>Length 259:</td>	Score 88:	DB 2:	Length 259:
Matches	26:	Conservative	12:	Mismatches	29:	Indels	24:
							Gaps
OY	5	HHHHHNGASGDLKNKVAOLKRVKSLKDKAELDKQ-----EVS	44				
Db	19	HHHHHNTPALGD--KRTVEYLKILISANDEQAESNRHFERHGILAVLMSSPGSKTT	76				
OY	45	RLNEIEDLK--AKTGLDINTSGIRRPAAKL	73				
Db	77	LLERTIELLKDELKIGTIEGDELTNRDAERI	107				
RESULT	6						
O9W6B1		PRELIMINARY:	324 AA.				
AC	O9W6B1:						
DT	01-NOV-1999 (TREMBlrel. 12, Created)						
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)						
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)						
DE	BASIC DOMAIN LEUCINE ZIPPER TRANSCRIPTION FACTOR.						
GN	MAP.						
OS	Brachydanio rerio (zebrafish) (Zebra danio).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Ostariophysi;						
OC	Cypriniformes; Cyprinidae; Rasbora; Danio.						
OX	NCBI_TaxID=7955;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	MEDLINE=99173794; PubMed=10072793;						
RX	Schwarzstein M., Kirn A., Hafler P., Cordes S.P.;						
RT	*Expression of Zkml2, a homologue of the Kml1/val segmentation gene,						
RL	during embryonic patterning of the zebrafish (Danio rerio).";						
DR	Mech. Dev. 80:223-226(1999).						
DR	EMBL: AF109781; AAD26141.1; .						
DR	InterPro: IPR001871; .						
DR	SMART: SM00338; BRL; .						
SO	SEQUENCE 324 AA; 36025 MW; C1ED46858C011EA8 CRC64;						
Query Match	23.1%	Score 87:	DB 13:	Length 324:			
Best Local	Similarity	24.2%	Pred. No. 4.5:				
Matches	32:	Conservative	11:	Mismatches	19:	Indels	70:
							Gaps
OY	5	HHHHHNG-----SMASGDL-----KN	20				
Db	176	HHHHHHHNOHQTPASRPSSTNSLSPTTRPRDRFSDQLVTMSVRLNHNHVGSKK	235				
OY	21	KVAOLKRVKSLKDKA-----ALKQEVSRLENEIEDLKA	55				
Db	236	EYIRLKKQRRRTLKNKGVAOSSRYKKRVQOHNILEKTKTOLLQNDHLKQEISRLVREDAVKE	295				
OY	56	KIGDLNNTSGIR	67				
Db	296	KYEKLVN--SGFR	306				
RESULT	7						
O31405		PRELIMINARY:	155 AA.				
AC	O31405:						
DT	01-NOV-1996 (TREMBlrel. 01, Created)						
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)						
DT	01-NOV-1998 (TREMBlrel. 08, Last annotation update)						
DE	MHC B-G ANTIGEN (FRAGMENT).						
OS	Gallus gallus (Chicken).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;						
OC	Gallus.						
OX	NCBI_TaxID=9031;						
RN	[1]						

```

RP SEQUENCE FROM N.A.
RX MEDLINE=90077532; PubMed=2592020;
RA Kaufman J., Salomonsen J., Skjodt K.;
RT "B-G CDNA clones have multiple small repeats and hybridize to both
RT chicken MHC regions.";
RL Immunogenetics 30:440-451(1989).
DR EMBL: M27668; AAA48624.1; -.
KW MHC.
FT NON_TER
SQ SEQUENCE 155 AA; 18201 MW; 7FC66EF2584F3978 CRC64;

Query Match 22.8%; Score 86; DB 7; Length 155;
Best Local Similarity 27.5%; Pred. No. 2.7; Matches 23; Indels 18; Gaps 2.
Matches 22; Conservative 17; Mismatches 23; Indels 18; Gaps 2.

QY 7 HHNHSASGDLKNKVA-----OLKKRVSLKDKAELKQEVSRLENEIED 52
   | : : : | | : | | : : : | | : : : | | : : : | | : : : |
Db 19 HCNEDDLSADLKKQIAELVEQREGVEEMHQLRKHYELKSSRAANLTKQLKLENEIE 78

QY 53 LKAKIGDLNNTSGIRPPAK 72
   : : : : : | | : |
Db 79 VEKHL----KKTGRRAPMLK 94

RESULT 8
042290 PRELIMINARY; PRT; 286 AA.
042290
AC 042290;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE BZIP TRANSCRIPTION FACTOR L-MAF.
GN L-MAF.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LENS;
RX MEDLINE=96192815; PubMed=9525857;
RA Ogino H., Yasuda K.;
RT "Induction of lens differentiation by activation of a bzip
RT transcription factor, L-Maf.";
RL Science 280:115-118(1998)
DR EMBL: AF034570; AAC15781.1; -.
DR InterPro; IPR001871; -.
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 286 AA; 32445 MW; 15DBD7B9B8DFA522 CRC64;

Query Match 22.7%; Score 85.5; DB 13; Length 286;
Best Local Similarity 25.9%; Pred. No. 5.4; Matches 25; Indels 49; Gaps 3.
Matches 30; Conservative 12; Mismatches 25; Indels 49; Gaps 3.

QY 4 SHHHNHG-----SMASGDL-----KNKVQLKRVKSLDK----- 35
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 AHHHHHHHLLLEERFSDQLYSMSVRELNROLRGFSKEEYIRLQONRTLLNKGTAOSC 216

QY 36 -----AAELKQEVSRLENEIEDLKAKIGDLNNTSGIRPPA 70
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 217 RYKRYQQRHILLENKCGQLQSQVEQLKQEVSRLLAKRBDLYKEKYELKLAARGFRRPS 272

RESULT 9
09H1F1 PRELIMINARY; PRT; 323 AA.
AC 09H1F1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

```

```

DE DJ644LL.1(KREISLER (MOUSE) MAP-RELATED LEUCINE ZIPPER HOMOLOG).
OS HOMO sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RL Ramsey H.;
DR Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AL035665; CAB75663.1; -.
SO SEQUENCE 323 AA; 35792 MW; A0FC09F8936CB16 CRC64;

OY Query Match 22.7%; Score 85.5; DB 4; Length 323;
Best Local Similarity 23.1%; Pred. No. 6;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;

OY 5 HHHHH-----GSMAS-----15
|||||
162 HHHHHHQAAPPSSAASAOQLPTSHPECPHATASATAAGNGSVEDREFSDQLVMSV 221
|||||

OY 16 -----GDLKNVVAOLKRRVRSIKDK-----AAELK 40
|:::|:::|:::|
222 RELNRHLNGFKFDEYIRLKQKRRTLLKNNGYAOSCRKKRVQOKHLENEKTOLIOVEQLK 281
|||||

OY 41 QEVSRLENEIEDLKAKIGDLNNTSGIR 67
|||||
DB 282 QEVSRLENERDAYKKCEKLAN-SGFR 307

RESULT 10
Q31620 PRELIMINARY; PRT; 321 AA.
AC Q31620;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE B-G (FRAGMENT).
GN B-G.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91239571; PubMed-1903541;
RA Miller M.M., Goto R., Young S., Chitivelva J., Hawke D., Miyada C.G.;
RT "Immunoglobulin variable-region-like domains of diverse sequence
within the major histocompatibility complex of the chicken."
RL Proc. Natl. Acad. Sci. U.S.A. 88:4377-4381(1991).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL: M61862; AAA48627.1; -.
DR InterPro: IPR003006; -.
DR InterPro: IPR003600; -.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00410; IG_Like; 1.
DR MHC.
FT NON_TER 1
SQ SEQUENCE 321 AA; 36360 MW; 7F25BD29E39860C7 CRC64;

OY Query Match 22.5%; Score 85; DB 7; Length 321;
Best Local Similarity 26.8%; Pred. No. 6;
Matches 22; Conservative 18; Mismatches 24; Indels 18; Gaps 2;

OY 7 HHHGSMASGDLK-----NKVAOLKRRVRSIKDKAAELKQEVSRLENEIED 52
|||||
DB 185 HLAERDLSADKLKLAALVDEQREAVEERDSOLRKQYEKLSRATWLKTLKLENEIDE 244
|||||

OY 53 LKAATGDLNNTSGIRPPAKLN 74

```

```

Db      245 VEKHLKRI---GIRAPNLIK 262
      : : : : : : : : : : : :
RESULT  11
Q9JLAX2
ID      Q9JLAX2      PRELIMINARY:      PRT:      479 AA.
AC      Q9JLAX2;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE      PSPA (FRAGMENT).
GN      PSPA.
OS      Streptococcus pneumoniae.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC      Streptococcus.
OX      NCBI_TaxID=1313;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BG7817;
RA      Hollingshead S.K., Becker R., Briles D.E.;
RT      "Diversity of Psppa: mosaic genes and evidence for past recombination
RT      in Streptococcus pneumoniae.";
RL      Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF071826; AAF27719.1; -.
DR      InterPro; IPR000533; -.
DR      InterPro; IPR002965; -.
DR      PRINTS; PR01217; PRICHEXTENSN.
DR      PRINTS; PRO0194; TROPOMOSIN.
FT      NON_TTR      479      479
SO      SEQUENCE      479 AA; 53257 MW; B9C0D2CA15DE3654 CRC64;
```

	Query Match	22.5%	Score 85;	DB 2;	Length 479;	
	Best Local Similarity	34.9%	Pred. No. 9.6;	Mismatches 13;	Indels 0;	Gaps 0
	Matched 15;	Conservative 15;	Mismatches 13;	Indels 0;	Gaps 0	
OY	17 DKKNRVAAQLKRRVRSLLDKMAELKOEVSRLNEIEDLAKICD 59					
	: :: ::: : :: :: : :: :					
DB	215 ELQNKVADLEKEIADADEKVADLEKEVALEKDVEEGFKSDGE 257					
RESULT 12						
O9LAX3						
ID O9LAX3:	PRELIMINARY;	PRT:	480 AA.			
AC O9LAX3:						
DT 01-OCT-2000 (TREMBLrel. 15, Created)						
DF 01-OCT-2000 (TREMBLrel. 15, Last sequence update)						
DI 01-MAR-2001 (TREMBLrel. 16, Last annotation update)						
DE PSPA (FRAGMENT).						
GN PSPA.						
OS Streptococcus pneumoniae.						
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;						
OC Streptococcus.						
OX NCBI_TaxID=1313;						
RN [1]						
RP SEQUENCE FROM N.A.						
RC STRAIN=BG7561;						
RA Hollingshead S.K., Becker R., Biles D.E.;						
RT "diversity of Pspa: mosaic genes and evidence for past recombination in Streptococcus pneumoniae."						
RL Submitted (JUN-1998) to the EMBL/Genbank/DDB databases.						
EMBL AF071824; AAF27718.1; -.						
DR InterPro: IPR000533; -.						
DR InterPro: IPR002965; -.						
DR PRINTS: PRO1217; PRICHEXTENS.						
DR PRINTS: PRO0194; TROPOMYCIN.						
FT NON_TER 480						
SQ SEQUENCE 480 AA: 53043 MW: DA01C9E0190D7A0 CRC64;						
Query Match	22.5%	Score 85;	DB 2;	Length 480;		
Best Local Similarity	34.9%	Pred. No. 9.7;				

Matches 15; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

OY 17 DLKKNVAQLKRVSLKDKAELKOEVSRLNEIEDLKAKIGD 59
 Db 215 ELONKVADELEKEIADAEKVADELEKEVAKLEKDEGEKRESDE 257

RESULT 13
 Q9UZC8

ID Q9UZC8 PRELIMINARY; PRT; 880 AA.

AC Q9UZC8;

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)

GN PAB0812.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.

NCBI_TaxID=29292;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-ORSAY;

RA Helig R.;

RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome

structure and evolution.";

Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

EMBL: AJ248286; CAB50131.1; -

DR InterPro: IPR000533; -

DR InterPro: IPR001238; -

DR InterPro: IPR001687; -

DR InterPro: IPR003439; -

DR InterPro: IPR003593; -

DR Pfam: PF00470; RecF; 1.

DR PRINTS: PR00194; TROPOMOSIN.

DR SMART: SM00382; AAA; 1.

SO SEQUENCE 880 AA; 103970 MW; FDB177EC7E026479 CRC64;

Query Match 22.5%; Score 85; DB 1; Length 880;

Best Local Similarity 37.2%; Pred. No. 17;

Matches 16; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

OY 18 LKKNVAQLKRVSLKDKAELKOEVSRLNEIEDLKAKIGD 60
 Db 240 INKISIELKIQVEKLGKRGKLEKIVIERSTIEKKAKISEL 282

RESULT 14
 Q9V217

ID Q9V217 PRELIMINARY; PRT; 281 AA.

AC Q9V217;

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE HYPOTHEICAL 32.7 KDA PROTEIN.

GN PAB2181.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.

NCBI_TaxID=29292;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-ORSAY;

RA Helig R.;

RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome

structure and evolution.";

Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

EMBL: AJ248283; CAB49181.1; -

DR InterPro: IPR000533; -

DR InterPro: IPR002017; -

DR PRINTS: PR00194; TROPOMOSIN.

DR Hypothetical protein

SO SEQUENCE 281 AA; 32744 MW; 8AA7BC80D1B5B877 CRC64;

Query Match 22.4%; Score 84.5; DB 1; Length 281;

Best Local Similarity 40.8%; Pred. No. 6.4;

Matches 20; Conservative 10; Mismatches 12; Indels 7; Gaps 1;

OY 17 DLKKNVAQLKRVSLKDKAELKOEVSRLNEIEDLKAKIG 58
 Db 222 ELERKVSLESLNEIETKVKSLKLEKKELEENKVKLEEEVKNLKIG 270

RESULT 15
 Q9Y503

ID Q9Y503 PRELIMINARY; PRT; 323 AA.

AC Q9Y503;

DT 01-NOV-1999 (TREMblrel. 12, Created)

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE MAFB/KREISLER BASIC REGION/LEUCINE ZIPPER TRANSCRIPTION FACTOR.

GN MAFB.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

NCBI_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE=BONE MARROW;

RA Wang P.W., Eisenbart J.D., Barsh G.S., Stoffel M., Le Beau M.M.;

"Human MAFB: cDNA cloning, genomic structure, and evaluation as a

candidate tumor suppressor gene in myeloid leukemias.";

Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

EMBL: AF134157; AAD30106.1; -

DR InterPro: IPR001871; -

DR SMART: SM00338; BRLZ; 1.

SO SEQUENCE 323 AA; 35829 MW; AE4DC23408E36E55 CRC64;

Query Match 22.4%; Score 84.5; DB 4; Length 323;

Best Local Similarity 23.1%; Pred. No. 7.3;

Matches 34; Conservative 8; Mismatches 20; Indels 85; Gaps 4;

OY 5 HHHHHH-----GSMAS-----
 Db 162 HHHHHHQAAPPSSAASPAOQLPTSHPGPHATASATAAGNGSVEDRSDDOLYSMSV 221

OY 16 -----GDLKKNVAQLKRVSLKDK-----AAELK 40
 Db 222 RELNRHLRGFTKDEVIRLKLKRRKLKNGYAOSCRKRYQOKHHLNENKQQLIQVEQLK 281

OY 41 QEVSRLENEIEDLKAKIGDINNTSGIR 67
 Db 282 QEVSRLENERDAYKVKCEKLAN-SGFR 307

Search completed: July 3, 2001, 15:05:43
 Job time: 608 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:13 ; Search time:86.32 Seconds
(Without Alignments)
239,490 Million cell updates/sec

Title: US-09-490-291-9
Perfect score: 1809
Sequence: 1 AEIYNKDKNVDLYGKAVGL.....NKLGVSDDTVAVGIYQFA 341

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	20	AAV42550
2	1792	99.1	340	17	AAW92998
3	1051	58.1	367	20	AAV42549
4	1051	58.1	367	20	AAV34058
5	1051	58.1	367	21	AAV57356
6	1029	56.9	377	20	AAV34057
7	1029	56.9	377	21	AAV57355
8	858.5	47.5	323	20	AAV34056
9	858.5	47.5	323	21	AAV57354
10	276	15.3	89	18	AAW27781
11	170.5	9.4	353	20	AAV26025

12	162.5	9.0	353	20	AAV26023	OMP protein of Pa
13	162	9.0	353	11	AAW06038	Class II outer mem
14	157	8.7	342	16	AAW70763	Mature class 2 por
15	157	8.7	342	18	AAW21742	Neisseria meningit
16	157	8.7	363	18	AAW70764	Fusion Class 2 por
17	157	8.7	363	18	AAW21743	Neisseria meningit
18	152.5	8.4	343	20	AAV26024	OMP protein of Pa
19	150.5	8.3	343	18	AAW21744	Neisseria meningit
20	147.5	8.2	309	16	AAW70762	Meningococcal grou
21	147.5	8.2	309	18	AAW21741	Neisseria meningit
22	144.5	8.0	341	11	AAW07043	P3 gene product of
23	140	7.7	372	16	AAW66607	Gonococcal porin-5
24	138.5	7.7	342	16	AAW66607	H. influenzae prot
25	138.5	7.7	342	16	AAW66680	H. influenzae prot
26	138.5	7.7	361	11	AAW05999	P2 gene product of
27	138.5	7.7	361	16	AAW66877	H. influenzae P2 e
28	138	7.6	390	18	AAW21678	Haemophilus influe
29	137.5	7.6	341	11	AAW07042	P3 gene product of
30	135.5	7.5	350	10	AAW90100	Protein IB of N. g
31	135.5	7.5	350	13	AAW27484	Protein IB. Synth
32	135.5	7.5	410	22	AAW19944	Moraxella catarrha
33	133	7.4	362	10	AAW90098	P2 antigen of Haem
34	131.5	7.3	363	16	AAW66878	PNV-2 fusion const
35	130	7.2	328	11	AAW60636	Class IB outer mem
36	121	6.7	1122	16	AAW64927	Cytadhesin protein
37	119.5	6.6	326	11	AAW06037	Class IA outer mem
38	118	6.5	730	21	AAW29582	Arabidopsis thalia
39	118	6.5	752	21	AAW29581	Arabidopsis thalia
40	116.5	6.4	185	19	AAW56320	Haemophilus paraga
41	114.5	6.3	707	19	AAW41523	Blood group antige
42	113	6.2	467	21	AAW16023	E. coli proliferat
43	111.5	6.2	373	11	AAW06035	Class I outer memb
44	111.5	6.2	934	17	AAW92289	Cycloisomaltooligo
45	111.5	6.2	972	17	AAW92288	Cycloisomaltooligo

ALIGNMENTS

RESULT	1	
AAV42550		
ID	AAV42550 standard; Protein; 362 AA.	
XX		
AC	AAV42550:	
XX		
DT	20-DEC-1999 (first entry)	
XX		
DE	E. coli wild-type ompF protein.	
KW	Bacteria; attenuation; deletion; mutant; vaccine; immune response;	
KW	Gram negative; infection; diarrhoea; food poisoning; typhoid;	
KW	salmonellosis; gonorrhoea; gastroenteritis; whooping cough.	
XX		
OS	Escherichia coli.	
XX		
PN	W09949026-A1.	
XX		
PD	30-SEP-1999.	
XX		
PF	25-MAR-1999; 99MO-GH00935.	
XX		
PR	25-MAR-1998; 98GB-0006449.	
XX		
PA	(PEPT-) PEPTIDE THERAPEUTICS LTD.	
XX		
PI	Chatfield SN;	
XX		
DR	WPI; 1999-580447/49.	
XX		
PT	N-PSDB; AA222887.	
XX		
PT	New attenuated bacteria useful as vaccines for protecting against	
XX	Infections -	

KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough.
 XX Escherichia coll.
 XX MO9949026-A1.
 XX 30-SEP-1999.
 XX 25-MAR-1999; 99MO-GB00935.
 XX 25-MAR-1998; 98GB-0006449.
 XX (PEPT-) PEPTIDE THERAPEUTICS LTD.
 XX Chalfield SN;
 XX WPI: 1999-580447/49.
 XX N-PSDB; AA222885.
 XX New attenuated bacteria useful as vaccines for protecting against
 PT infections -
 PS Disclosure; Page 57-58; 69pp; English.
 XX This sequence represents the E. coli wild-type ompC protein. The coding
 CC sequence of the ompC gene was removed via PCR using primers TT7-TT10
 CC (AA222893-222896) to produce a non-reverting deletion mutation. The
 CC mutant ompC gene (AA222886) was used in the production of a bacterium
 CC attenuated by a non-reverting mutation in each of the ompC gene, the arcC
 CC gene (AA222883, AA222884), and the ompF gene (AA222887, AA222888). The
 CC mutant bacteria provide immunogenic activity with reduced virulence and
 CC thus can be used as a vaccine for raising an immune response against a
 CC variety of bacteria in a mammalian host. Such vaccines can provide
 CC protection against e.g., E. coli (a cause of diarrhoea in humans),
 CC Salmonella typhimurium (the cause of salmonellosis in several animal
 CC species), S. typhi (the cause of human typhoid), S. enteritidis (a cause
 CC of food poisoning in humans), S. choleraesuis (a cause of salmonellosis
 CC in pigs), S. dublin (a cause of both a systemic and diarrhoeal disease in
 CC cattle, especially of new-born calves), Haemophilus influenzae (a cause
 CC of meningitis), Neisseria gonorrhoeae (a cause of gonorrhoea), Yersinia
 CC enterocolitica (the cause of a spectrum of disease in humans ranging
 CC from gastroenteritis to fatal septicemic disease), Bordetella pertussis
 CC (the cause of whooping cough), and Brucella abortus (a cause of abortion
 CC and infertility in cattle and a condition known as undulant fever in
 CC humans).
 CC Sequence 367 AA;
 XX
 XX
 Query Match 58.1%; Score 1051; DB 20; Length 367;
 Best Local Similarity 58.9%; Pred. No. 9.6e-83;
 Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
 OY 1 AEIYNKDGNDVLDYKRAVGLHYFSKNGENSYGNGDMTYARLGFKEPTQINSDLTGCGQ 60
 DB 22 aeyynkdgndldygykvdglnhts-----dnkdvdgqcltymrlgfkgetytdqltgy9g 76
 OY 61 WEYFQGNNSGADAOGTGNKTRLAFLAKLYADVGSFDYGRNRYGVVYDALGYTDLMPERGQ 120
 DB 77 weyfqgnnsaen---emswtrvafaglkfdygsfdygrnrygvvydvtswtdlpefgy 133
 OY 121 DTVASDFFVGRVGVATYRNSNFFGLVDGLNFAVOYLGNKNERDT-----A 166
 DB 134 dtygsdfffmgqrfngfatyrntdftglvdglnfavgygqknpsgsgftsgvtngrda 193
 OY 167 RRSNGDGVGSISYEVGFGVGVAGVADRTNLOE-AQPLGNGKKAQDMAGLTYDANNI 225
 DB 194 lrrqgdgvgsisltydyegfglgaissskrtcdagntaaylmgdraeyleyglkydanni 253
 OY 226 YLAANYEETRNATITNKTFTSGFANKTODVLLVAQYQDFGFRPSIAYKSKAKDV-E 284
 DB 254 ylaaeyqctynatrv-----gslygwankagnfseavagqfifglrpsalaylqsgkynlgr 308

OY 285 GIGDVLNVEFGATYENKMSGYVDYIINOISDNKL-----GVSGDDTVAVGIYQF 340
 DB 309 gyddedllkyvdvagalyfnkmsltydykxlnlld-dnqftrdaglnclnvalaylyyqf 367
 RESULT 4
 ID AAY34058 standard; protein: 367 AA.
 AC AAY34058;
 XX 23-NOV-1999 (first entry)
 DE E. coli outer membrane protein c precursor.
 XX
 XX ulcerative colitis; histone; H1-like antigen; porin antigen;
 KW Bacteroides antigen; IBD; PANCA; inflammatory bowel disease;
 KW diagnosis; perinuclear anti-neutrophil cytoplasmic antibody;
 KW outer membrane protein c precursor.
 XX
 XX Escherichia coll.
 XX OS
 XX MO9945955-A1.
 XX 16-SEP-1999.
 XX 12-MAR-1999; 99MO-US05492.
 XX 12-MAR-1998; 98US-0041889.
 XX (REGC) UNIV CALIFORNTA.
 XX
 XX Braun J, Cohavy O;
 XX WPI: 1999-551215/46.
 XX
 PT Use of histone H1, porin or Bacteroides antigens as targets for the
 PT diagnosis, prevention and treatment of ulcerative colitis -
 XX
 XX Claim 8; Fig 11; 134pp; English.
 XX
 CC The invention provides a method for the diagnosis, prevention and
 CC treatment of ulcerative colitis (UC) using histone H1-like antigen, a
 CC porin antigen or a Bacteroides antigen as a target antigen. The novel
 CC method of diagnosing UC in a subject suspected of having inflammatory
 CC bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
 CC (2) contacting the sample with a histone H1-like antigen, or perinuclear
 CC anti-neutrophil cytoplasmic antibody (PANCA)-reactive fragment, to form a
 CC complex of the histone H1-like antigen, or the PANCA-reactive fragment,
 CC and antibody to the histone H1-like antigen; and (3) detecting the
 CC presence or absence of the complex; where the presence of the complex
 CC indicates that the subject has UC. The PANCA-reactive histone H1-like
 CC antigen, porin antigen and Bacteroides antigen are useful in the
 CC diagnosis, prevention and treatment of UC. The methods can also be used
 CC for identifying agents useful for treating UC. The present sequence
 CC represents a E. coli outer membrane protein c precursor.
 CC
 XX
 XX Sequence 367 AA;
 XX
 Query Match 58.1%; Score 1051; DB 20; Length 367;
 Best Local Similarity 58.9%; Pred. No. 9.6e-83;
 Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
 OY 1 AEIYNKDGNDVLDYKRAVGLHYFSKNGENSYGNGDMTYARLGFKEPTQINSDLTGCGQ 60
 DB 22 aeyynkdgndldygykvdglnhts-----dnkdvdgqcltymrlgfkgetytdqltgy9g 76
 OY 61 WEYFQGNNSGADAOGTGNKTRLAFLAKLYADVGSFDYGRNRYGVVYDALGYTDLMPERGQ 120
 DB 77 weyfqgnnsaen---emswtrvafaglkfdygsfdygrnrygvvydvtswtdlpefgy 133
 OY 121 DTVASDFFVGRVGVATYRNSNFFGLVDGLNFAVOYLGNKNERDT-----A 166


```
XX DR WPI; 1999-551215/46.
XX
XX PT Use of histone H1, porin or Bacteroides antigens as targets for the
XX diagnosis, prevention and treatment of ulcerative colitis
XX PS Claim 8; Fig 11; 134pp; English.
XX
XX The invention provides a method for the diagnosis, prevention and
XX treatment of ulcerative colitis (UC) using histone H1-like antigen, a
XX porin antigen or a Bacteroides antigen as a target antigen. The novel
XX method of diagnosing UC in a subject suspected of having inflammatory
XX bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
XX (2) contacting the sample with a histone H1-like antigen, or perinuclear
XX anti-neutrophil cytoplasmic antibody (PANCA)-reactive fragment, to form a
XX complex of the histone H1-like antigen, or the PANCA-reactive fragment,
XX and antibody to the histone H1-like antigen; and (3) detecting the
XX presence or absence of the complex; where the presence of the complex
XX indicates that the subject has UC. The PANCA-reactive histone H1-like
XX antigen, porin antigen and Bacteroides antigen are useful in the
XX diagnosis, prevention and treatment of UC. The methods can also be used
XX for identifying agents useful for treating UC. The present sequence
XX represents a E. coli outer membrane protein F precursor.
XX
XX Sequence 323 AA;
XX
XX Query Match 47.5%; Score 858.5; DB 20; Length 323;
XX Best Local Similarity 57.2%; Pred. No. 3.2e-66;
XX Matches 174; Conservative 33; Mismatches 72; Indels 25; Gaps 7;
XX
XX QY 1 AEYKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTVARLGFKEGTQINSDLTGYGQ 60
XX DB 22 aeaynkdknklidlygkvqglhyfsdnak-----dgdqsyarllfgkgetqindqltgygq 76
XX QY 61 WEYNQGNNSGADAGQCNKTRLAFLAGLKYADVGSFDYGRNYGVYDALGYTDMLEPFGG 120
XX DB 77 weynqgnnsesknqsw--rlaflaglkfadygsfdygrnygwydmleqvtcmlpefgy 134
XX QY 121 DT-AVSDPEFYGRVGVATYRNSNFFGLVDGLNFAVOYLGRNE-----RDTAR 167
XX DB 135 dytlnadnfmgrangvatlyrntdfjgvlnglnfavqyqgnnegasngqegtlngrd--vr 193
XX QY 168 RNSGDGVGSGISYEX-EGFGIVGAYGAADRNLQEAOPLGNGKKAQOMATGLKYDANNIY 226
XX DB 194 hengdgwglstctydlgmfsagaaytsedrindqynhtaaggdkadavtaglkydannily 253
XX QY 227 LAANGETRNAPTPTNKFTNTSGFANKTQDVLVAQYQFDGLRPSIATYTSKAKDVEGI 286
XX DB 254 latmysetrmtlpfgd---sdyavanktcqnftevtagydfdgllrpavsflnmskgydlhaa 310
XX
XX QY 287 GDVD 290
XX DB 311 gpad 314
XX
XX RESULT 9
XX ID AAY57354
XX AC AAY57354;
XX
XX 13-JUN-2000 (first entry)
XX
XX E. coli outer membrane protein F precursor.
XX
XX KW Ulcerative colitis; inflammatory bowel disease; porin antigen; MAAb;
XX PANCA; perinuclear anti-neutrophil cytoplasmic antibody;
XX histone H1; outer membrane protein F precursor.
XX
XX OS Escherichia coli.
XX
XX US6033864-A.
XX
XX PN
```

```
XX PD 07-MAR-2000.
XX
XX PF 12-MAR-1998; 9805-0041889.
XX
XX PR 12-APR-1996; 9605-0057846.
XX PR 11-APR-1997; 9705-0837058.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX PI Cohavy O, Braun J;
XX
XX DR WPI; 2000-255695/22.
XX
XX PT Diagnosing ulcerative colitis or susceptibility, by detecting complex
XX formation between microbial porin antigen and perinuclear
XX anti-neutrophil cytoplasmic autoantibodies
XX
XX PS Claim 1; Fig 10; 49pp; English.
XX
XX The invention provides a method for diagnosing ulcerative colitis in a
XX subject suspected of having inflammatory bowel disease. The method
XX comprises reacting a patient sample with a porin antigen that is
XX immunologically reactive with PANCA (perinuclear anti-neutrophil
XX cytoplasmic antibodies) and detecting formation of a Ag-PANCA complex
XX as indicative of ulcerative colitis. The method is used to diagnose
XX ulcerative colitis or susceptibility to it. The present sequence
XX represents a E. coli outer membrane protein F precursor.
XX
XX Sequence 323 AA;
XX
XX Query Match 47.5%; Score 858.5; DB 21; Length 323;
XX Best Local Similarity 57.2%; Pred. No. 3.2e-66;
XX Matches 174; Conservative 33; Mismatches 72; Indels 25; Gaps 7;
XX
XX QY 1 AEYKDGKNDLYGKAVGLHYFSKNGENSYGGNGDMTVARLGFKEGTQINSDLTGYGQ 60
XX DB 22 aeaynkdknklidlygkvqglhyfsdnak-----dgdqsyarllfgkgetqindqltgygq 76
XX QY 61 WEYNQGNNSGADAGQCNKTRLAFLAGLKYADVGSFDYGRNYGVYDALGYTDMLEPFGG 120
XX DB 77 weynqgnnsesknqsw--rlaflaglkfadygsfdygrnygwydmleqvtcmlpefgy 134
XX QY 121 DT-AVSDPEFYGRVGVATYRNSNFFGLVDGLNFAVOYLGRNE-----RDTAR 167
XX DB 135 dytlnadnfmgrangvatlyrntdfjgvlnglnfavqyqgnnegasngqegtlngrd--vr 193
XX QY 168 RNSGDGVGSGISYEX-EGFGIVGAYGAADRNLQEAOPLGNGKKAQOMATGLKYDANNIY 226
XX DB 194 hengdgwglstctydlgmfsagaaytsedrindqynhtaaggdkadavtaglkydannily 253
XX QY 227 LAANGETRNAPTPTNKFTNTSGFANKTQDVLVAQYQFDGLRPSIATYTSKAKDVEGI 286
XX DB 254 latmysetrmtlpfgd---sdyavanktcqnftevtagydfdgllrpavsflnmskgydlhaa 310
XX
XX QY 287 GDVD 290
XX DB 311 gpad 314
XX
XX RESULT 10
XX ID AAM27781
XX AC AAM27781;
XX
XX 21-JUL-1998 (first entry)
XX
XX Amino acid sequence of an outer membrane protein F precursor.
XX
XX KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
XX Staphylococcal gene; regulatory element; bacterial gene expression;
XX
```

Query Match	Best Local Similarity	Score	DB	Length
Matches 85; Conservative 41; Mismatches 132; Indels 109; Gaps 17;	9.4%;	170.5;	20;	353;
1 AEIYKDKGKVDLVGKAVGLHYFSKNGSGNSYGGNGDMTY--ARLGFKGEMQINSDLNGY 58	23.2%;	97.7e-07;		
21 aavynnegtkvelggrvstiaeqscsnkdkqkhpqslnggsrlnkvtlmhngdyal 80	41;	132;	109;	17;
59 GOMEYNFOGNNSEGADAOTGN-----KTRLAEPAGLKYADVGSFDDG----- 99				
81 gyyerelinkdldgnehkngsgfsgitcklayagngkelgatqlckltadkstaed 140				
100 RNYGVVYDALGYTDMLEPERGGDTAYSDDEPFGVGRGVGATYRNSNFFGLVDGLNFAVQILG 159				
141 keyaylek-----nsyilepnagalay-----lykg-----legltlgaasyv- 176				
160 KNERPTARNSNDGVGSGSYSEEFSGIVGAVGADRNRLDQAOLGNGKKAKEQATGLK 219				
177 -----fgngfnfsdye-----itdgkvsnavqvgak 201				
220 YDANNIYLAANYGETRNATPTINFKFTNTSGFANKTQDV---LTVAQYOFO-PFGLRPSI-- 273				
202 ydammivagfaygr-----tnykaqqakcgqynagalatqyhnfdldqllisdis 250				
274 AYTKSRKADVEGIGVDVLVNTF-EVGATYYFNKKNSTY--VDY-IINOISDNKLGVGSD 329				
251 gyakctknk-----adkhekryfvsppgfygelmednlgnlkyerlnsvdgggkcv---re 302				
330 DTVANGI 336				
303 havlfgl 309				

Haemophilus influenzae.
 WO9929724-A2.
 17-JUN-1999.
 08-DEC-1998; 98MO-US25990.
 21-JUL-1998; 98US-0120051.
 08-DEC-1997; 97US-0067957.
 (UYGE-) UNIV GEORGIA RES FOUND INC.
 Glisson JR, Luo Y;
 WPI; 1999-385575/32.
 DNA encoding Pasteurella multocida outer membrane protein H, useful for delivering peptide epitopes for vaccines against fowl cholera
 Disclosure: Fig 5; 11pp; English.
 The present amino acid sequence is that of the Haemophilus influenzae porin P2. This protein is a bacterial porin, which is a major outer membrane protein that is involved in pore formation. They serve as molecular sieves that allow polar solutes to pass through, but excludes non-polar molecules of comparable sizes. This amino acid sequence shares a sequence homology of about 38% to Pasteurella multocida x-73 OmpH protein. Peptides derived from the P multocida OmpH protein are useful in the production of protective vaccines for use in vertebrates in particular, for prevention of fowl cholera and for diagnosis in veterinary medicine.
 Sequence 353 AA;
 RESULT 12
 AA126023
 AA126023 standard; Protein: 353 AA.

XX	AAV26023:	
AC		
XX	29-SEP-1999	(first entry)
DT		
XX		
DE	Omph protein of Pasteurella multocida X-73.	
XX		
KW	Outer membrane protein H; omph gene; fowl cholera; molecular sieve;	
KW	Pasteurella multocida X-73; porin; avian; vaccine; veterinary medicine;	
KW	bacterial porin; H. influenzae porin P2.	
XX		
OS	Pasteurella multocida.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..20
FT		/label= Signal_peptide
FT	Protein	/note= "Hydrophobic amino acids"
FT		21..353
FT		/note= "mature Omph protein"
FT	Cleavage-site	7..9
FT		/note= "Present within the signal peptide"
FT	Cleavage-site	9..11
FT		/note= "Present within the signal peptide"
FT	Cleavage-site	12..14
FT		/note= "Present within the signal peptide"
FT	Cleavage-site	18..20
FT		/note= "Present within the signal peptide"
FT	Peptide	81..110
FT		/label= MAP-I2 peptide
FT		/note= "Multiple Antigen Peptide derived from loop 2"
FT	Peptide	217..238
FT		/label= MAP-I5_peptide
FT		/note= "Multiple Antigen Peptide derived from loop 5"
PN	W09929724-A2.	
XX		
PD	17-JUN-1999.	
XX		
PE	08-DEC-1998;	98WO-US25990.
XX		
PR	21-JUL-1998;	98US-0120051.
PR	08-DEC-1997;	97US-0067957.
XX		
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.	
XX		
PI	Glisson JR, Luo Y;	
DR	WPI: 1999-385575/32.	
DR	N-PSDB: AAX80609.	
XX		
PT	DNA encoding Pasteurella multocida outer membrane protein H, useful	
PT	for delivering peptide epitopes for vaccines against fowl cholera	
XX		
PS	Claim 10; Fig 4; 119pp; English.	
XX		
CC	The present amino acid sequence is that of the avian Pasteurella	
CC	multocida X-73, outer membrane protein H (Omph). This protein is a	
CC	porin, which is a major outer membrane protein that is involved in	
CC	pore formation. They serve as molecular sieves that allow polar solutes	
CC	to pass through, but excludes non-polar molecules of comparable sizes.	
CC	The amino acid composition is typical of non-specific bacterial porins	
CC	with highly negative hydropathy index, high glycine content, low proline	
CC	content and lack of cysteine. This amino acid sequence shows homology of	
CC	about 38% to bacterial porins, like that of H. influenzae porin P2.	
CC	Multiple antigen peptides are synthesized from the loop 2 and loop 5	
CC	regions of the protein that can be used for immunisation of the fowls	
CC	to induce heterologous protection against P. multocida infection. This	
CC	sequence can be used for diagnostic assays for use in veterinary	
CC	medicine, especially for the diagnosis of fowl cholera, caused by	
CC	P. multocida. Peptides derived from the Omph protein are useful in the	
CC	production of protective vaccines for use in vertebrates, in particular	
CC	for prevention of fowl cholera.	
XX		

Seq	Sequence	353 AA;
Query Match	9.08; Score 162.5; DB 20; Length 353;	
Best Local Similarity	22.94; Pted. No. 3.8e-06;	
Matches	83; Conservative 39; Mismatches 150; Indels 91; Gaps	
Oy	1 AEINRDKGNKVDLGGKAVGLHYFSKNGENSYGGNGDMFYARLGFKGETQINSDLNGYQ 60	
Db	21 atyngdgtkxdvng-slrlllkkheknrgdlvng-----srysfkashlgeglasalay 75	
Oy	61 WEYNFGNNGSEGADATGKNKTR-----LAFAGLKADYGSFDYGRNCGVT 106	
Db	76 telftsknvpyqvkdgqgevvreyeveklgnvnhvkrlyagfayeglygtlftgnqltl-- 133	
Oy	107 DALGTDLPEFGGDDTAYSDDFFGRRGVATYRNSNPFGLVNGSLNFAVGYLCKNERDFA 166	
Db	134 -----gddvglsd-----ykyfnsglmmll----- 153	
Oy	167 RRSNGDVGSGISYEYEGFIVGAY---GAADRNLQEAQ-----PLNGKKAEQWATGL 218	
Db	154 --ssgekalnfkasefnfgltggayvtsadadkqalrdgrgfvaglynnkmgd---vgf 208	
Oy	219 KYDANNITLANY--GETRNRATPTNKFNTSGFRANKTQDVLVAQYQDFGGLRPSIAT 276	
Db	209 afeag---ysqkyvkqveqgnppaaqkv----fkdekekaflmgaelsyaglaigydy 260	
Oy	277 KSKAKDVEGIDVDLVNVEFGATYRFNKNMSTYVDYIIQIDSDNKLGVSGDDTVAVG 336	
Db	261 gskvtvndg-----ktralevglnydlndrakvytdfwek--egpkgdvtrnrtvavgf 313	
Oy	337 VYQ 339	
Db	314 gylk 316	
RESULT 13		
AA06038		
ID	AA06038 standard; protein; 353 AA.	
XX		
AC	AA06038;	
XX		
CT	20-NOV-1990 (first entry)	
DE	Class II outer membrane protein of N.meningitidis.	
XX		
KW	Meningococcal disease; meningitis; vaccine;	
KW	Class I outer membrane protein; ds.	
XX		
OS	Neisseria meningitidis.	
XX		
PN	W09006696-A.	
PD	28-JUN-1990.	
PF	19-DEC-1989; 89WC-US05678.	
XX		
PR	26-JUN-1989; 89NL-0001612.	
PR	19-DEC-1988; 88NL-0003111.	
PR	06-JAN-1989; 89NL-0000030.	
XX		
PA	(PRAX-) PRAXIS BIOLOGICS IN.	
PA	(VOLK-) RIJCKINSINST VOLKSGEZONDH.	
XX		
PI	SEID RC, PARADISO PR, POOLMAN JT, HOOGERHOUT P, WIERTZ EJ;	
XX		
PI	VANDERLEY P, HECKELS JE, CLARKE IN;	
XX		
DR	WPI: 1990-224326/29.	
XX		
PT	Meningococcus class I outer-membrane protein vaccine - useful to	
XX	immunise against meningococcal disease.	

PS Disclosure: ; 120pp; English.

XX Peptides derived from outer membrane, encoding homo/heterologous
CC product such as an antigen-flagelin fusion protein, are useful in
CC eliciting a strong and wide ranging immune response against most
CC serotypes.

XX Sequence 353 AA:

SO

Query Match 9.0%; Score 162; DB 11; Length 353;
Best Local Similarity 26.6%; Pred. No. 4.2e-06;
Matches 79; Conservative 42; Mismatches 126; Indels 50; Gaps 16;

QY 41 ARLGFKGTQINSDLTGYGWEYTFQGNSEGADQTKTRLAFLAKYADGSPDYGR 100
DB 54 sklgfsgqdelglnkaiwqlqle---qksiaigtngswgn---rgsfllgk---ggfgtivr 104

QY 101 --NYGVVYDALGYTDMPEFGDPTAVSDDFYGVGVGATVYRNPNFG--LVQGLNFAVQ 156
DB 105 aglnlntvlykagadnvnawesgsnt--edvlgltgrvesrelsvrldpvtfagfsgsvq 162

QY 157 YLKG---NERDTPARRSNGDVGSGISYEYEGFGIVGAYGA-----ADR--TNLQ 200
DB 163 yperdanndvdkykhktsaresyhaqlkyenagffgqyagsfakyaadlnndaervavnta 222

QY 201 EAQPLNGKKAEGQWANGTKYDANNIY--LAANYGETRNATPTTNKFTNTSGFANKTQDVL 258
DB 223 naphv---kdyqvhrrvavagydandlysvavagyeaakn-----nevgsstkghkhtq-va 273

QY 259 LVNQYQDFGLRPSIATKSKAKDVEGIDVDLVNFE---VCATYYFFNKMSTYV 311
DB 274 ataayrtfg-nvcpvryahgfkakvngvkd---anyqddqvivgadydfskrtsalv 326

RESULT 14
AAR70763
ID AAR70763 standard; Protein: 342 AA.

XX AAR70763;

XX 24-AUG-1995 (first entry)

XX Mature class 2 porin expressed from gene cloned into pET-17b.

XX Group B porin protein; meningococcal; outer membrane protein;
KW vaccine; meningitis; Neisseria meningitidis; PCR primer.

XX Synthetic.

XX WO9503413-A.

XX 02-FEB-1995.

XX 22-JUL-1994; 94WO-0508327.

XX 23-JUL-1993; 93US-0096182.

XX (NAVA-) NORTH AMERICAN VACCINE INC.
PA (UYRO) UNIV ROCKEFELLER.

XX Blake MS, Hronowski LJ, Liang S, Pullen JK, Qi HL;
PI Tal JY;

XX WPI: 1995-075239/10.
DR N-PSDB: AAO85391.

XX High expression of outer membrane meningococcal group B porin
PT proteins - and fusion proteins in Escherichia coli, and
PT purification method; for use in vaccines against Neisseria
PT meningitidis and in research.

XX Example: Figure 9; 81pp; English.

XX The example concerns cloning of Class 2 porin from Group B
CC Neisseria meningitidis strain BNCV M986 serotype 2a (see
CC AA085300, AA085301). The plasmid pET-17b was used to express the
CC class 2 porin. A plasmid was designed to yield a mature class 2
CC porin. The mature class 2 porin was constructed by amplifying
CC the pUC19-class 2 porin construct using the oligos AA085302 and
CC AA085303. This strategy allowed the cloning of the amplified class 2
CC porin into the NdeI and XhoI sites of the plasmid pET-17b thus
CC producing mature class 2 porin. Std. PCR was conducted using the
CC pUC19-class as the template and the two oligos. The PCR reaction
CC yielded a 1.1kb product. The DNA obtd. was purified and digested
CC with NdeI and XhoI and ligated to pET-17b, and used to transform
CC E.coli. DNA from transformed clones was sequenced. AA085291/R70763
CC show the nt and translated AA sequence of the mature class II
CC porin gene cloned into the expression plasmid pET-17b.

XX Sequence 342 AA:

SO

Query Match 8.7%; Score 157; DB 16; Length 342;
Best Local Similarity 24.7%; Pred. No. 1.1e-05;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLTGKAVGLHFFKNGCENSYGNGDMTY-----ARLGFKGTQINSDLTGYGQW 61
DB 3 vlygtlkagvevsvrxdagtykagqgsktatqadfgsklgtkqgedlgnmkaiwql 62

QY 62 EYNQGNNSSEGADQTKTRLAFLAKYADVGSFDYGR--NYGVVYDALGYTDMPEFG 119
DB 63 e---qksiaigtngswgn--rgsfllgk---ggfgrvraenlntvlyksgdnvnawesg 113

QY 120 GDTAVSDDFYGVGVGATVYRNPNFG--LVQGLNFAVQYLGK---NERDTPARRSNGDGV 174
DB 114 snt--edvlgltgrvesrelsvrldpvtfagfsgsvqyprdanndvdkykhktsare 171

QY 175 GGSISYEYEGFGIVGAYGA-----ADR--TNQEAQPLNGKKAQWMTGLKYD 221
DB 172 syhaglkynagffgqyagsfakyaadlnndaervavntaahp---kdyqvhrrvavagd 228

QY 222 ANNIIY--LAANYGETRNATPTTNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSK 279
DB 229 andlysvavagyeaakn-----nevgsstkghkhtqyaataayrtfg-nvcpvryahgf 282

QY 280 AKDVEGIDVDLVNFE---FEVGATYYFFNKMSTYV 311
DB 283 kakvngvkd---anyqddqvivgadydfskrtsalv 315

RESULT 15
AAW21742
ID AAW21742 standard; Protein: 342 AA.

XX AAW21742;

XX 09-MAR-1998 (first entry)

XX Neisseria meningitidis class 3 mature porin protein.

XX Porin protein; porB gene; outer membrane protein; M83;
KW meningitis; vaccine.

XX Neisseria meningitidis serogroup B strain 8765 (B:15:PL.3).
OS

XX WO9728273-A1.

XX 07-AUG-1997.

XX 31-JAN-1997; 97WO-0501687.

XX 13-JUN-1996; 96US-0020440.

XX 01-FEB-1996; 96US-0010972.

PA (NAVA-) NORTH AMERICAN VACCINE INC.

PI	Donets M, Liang S, Michon F, Minetti CA, Polvino-Bodnar M,
PI	Tal JY, Wang M, Minetti CASA;

Tai JY, Wang M, Minetti CASA,

DR WPI; 1997-402625/37.
DR N-PSDB; AAT77815-16.

N-PSDB; AA177815-16.

PT Production of meningococcal group B porin proteins in yeast - by
PT extraction of a porin protein, or porin protein fused to a yeast
PT secretion signal peptide

secretion signal peptide

PS Example 3; Fig 13A; 138pp; English.

CC This prein comprises the mature class 3 porin (M33) protein porB
CC from group B *Neisseria meningitidis*, encoded by a mature class 3
CC porin gene (see AAT77815). In vector pET-24a and expressed in *E. coli*
CC DH5-alpha transformants. A method is claimed for the high level
CC expression of an outer membrane meningococcal group B porin protein
CC in yeast. It comprises: ligating a DNA sequence encoding mature
CC porin (see AAT77815), or mature porin fused to a secretion signal
CC sequence (see AAT77814), into a vector containing a selectable marker
CC and a yeast promoter such as the AOX1 promoter; using the vector to
CC transform yeast, preferably *Pichia pastoris*, host cells; and
CC inducing expression of the recombinant protein. A method is also
CC claimed for purifying the porin protein. The level of porin
CC protein expression is enhanced by optimising the 5' region of the
CC porin protein gene for yeast codon usage (see AAT76738 and AAT77816).
CC The products are used to provide an immune response in mammals to
CC *N. meningitidis* (claimed).

N. meningitidis (claimed).

N. meningitidis (claimed).

N. meningitidis (claimed).

N. meningitidis (claimed).

N. meningitidis (claimed).

N. meningitidis (claimed).

N. meningitidis (claimed).

N. meningitidis (claimed):

SQ Sequence 342 AA;

Query Match	Score	DB	Length
8.78;	157;	18;	342;

Best Local Similarity 24.7%; Pred. No. 1,1e-05;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

83; Conservative 45; Mismatches 150; Indels 38; Gaps 10

```

Oy      11 vDIyKKAvgLHfEFGKNGENySNGNGMTy-----ARLEFGKGTQJNSOLTGyGw 61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3 vLjygtLkaYgvevsvkdagYyKkaqgkskstaqtadfgsklfgKkggedLgngmkatwJ 62

Oy      62 EYnFGGNSECADQOTGKNKTPLAFAGLKYADVSFdyGr--nGcYvYDALGyDMLPEFG 119
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      63 e---qkaSlagtnagwv--qgsfIgtL---ggfGtGvagnIntlvkksdghvnaWesg 113

Oy      120 GDTATSDPEFGPRGAGVATYRNSNPFg--LVDGnGFNVQYLGK--NEBRTARsNGDv 174
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      114 ntc--edvlgYtGtgrvesrsvrpadvfaGfsGsvgyprcdhndvdkYkhtkres 171

Oy      175 GGSISYETEGEGYVGATyA-----ADr--TnLEQnOPLGnKKKdQnATGtKYD 221
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      172 syhaGklyeengfIfGyqygsfakYadIntcdereavntaahpy--kdyqvhvnaYd 228

Oy      222 ANNIY--LAANYGFERNATPTfTnKfTNTSGFRANKTODVLVAOYQDFGRLPSIAYtRSK 279
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      229 andlysvsvaggyaakn-----nevgsctkYgkKheqtqyaataayrG-nvLprvsvyhgF 282

Oy      280 AKDVGEGIDvDLvNv-----FEVgATYTYRnKKMSYv 311
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      283 kaKvngvkd---aYqyqdvIvgyadYdsktsalv 315

```

283 kakvngvkd---anyqdv1vgadyatskrtsalv 315

315 vgaadyatskrtsaiv

283 kakvngvkd---anyqdv1vgadyatskrtsalv 315

315 vgaadyatskrtsaiv

283 kakvngvkd---anyqdv1vgadyatskrtsalv 315

315 vgaadyatskrtsaiv

283 kakvngvkd---anyqdv1vgadyatskrtsalv 315

315 vgaadyatskrtsaiv

283 kakvngvkd---anyqdv1vgadyatskrtsalv 315

315 vgaadyatskrtsaiv

283 kakvngvkd---anyqdv1vgadyatskrtsalv 315

```
Search completed: July 3, 2001, 14:58:14
Job time: 444 sec
```

Job time: 444 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:45 ; Search time 48.42 Seconds
(without alignments)
141.869 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1 AEIYNKGDKNVDLYGKANGL.....NKLGVGSDDTVAVGIVQFA 341

Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.1	340	2	US-08-355-844-1 Sequence 1, Appl1
2	1792	99.1	340	5	PCT-US95-16126-1 Sequence 1, Appl1
3	1051	58.1	367	3	US-09-041-889-30 Sequence 30, Appl1
4	1029	56.9	377	3	US-09-041-889-29 Sequence 29, Appl1
5	858.5	47.5	323	3	US-09-041-889-28 Sequence 28, Appl1
6	157	8.7	342	1	US-08-096-182A-4 Sequence 4, Appl1
7	157	8.7	342	1	US-08-096-182A-4 Sequence 4, Appl1
8	157	8.7	342	3	US-08-798-760-4 Sequence 4, Appl1
9	157	8.7	342	5	PCT-US94-08327-4 Sequence 4, Appl1
10	157	8.7	363	1	US-08-096-182A-6 Sequence 6, Appl1
11	157	8.7	363	1	US-08-877-109-6 Sequence 6, Appl1
12	157	8.7	363	3	US-08-798-760-6 Sequence 6, Appl1
13	157	8.7	363	5	PCT-US94-08327-6 Sequence 6, Appl1
14	147.5	8.2	309	1	US-08-096-182A-2 Sequence 2, Appl1
15	147.5	8.2	309	1	US-08-877-109-2 Sequence 2, Appl1
16	147.5	8.2	309	3	US-08-798-760-2 Sequence 2, Appl1
17	147.5	8.2	309	5	PCT-US94-08327-2 Sequence 2, Appl1
18	138.5	7.7	342	4	US-08-096-181A-12 Sequence 12, Appl1
19	138.5	7.7	342	4	US-08-096-181A-14 Sequence 14, Appl1
20	138.5	7.7	342	5	PCT-US94-08326-12 Sequence 12, Appl1
21	138.5	7.7	342	5	PCT-US94-08326-14 Sequence 14, Appl1
22	138.5	7.7	361	4	US-08-096-181A-8 Sequence 8, Appl1
23	138.5	7.7	361	4	US-08-096-181A-10 Sequence 10, Appl1
24	138.5	7.7	363	5	PCT-US94-08326-10 Sequence 10, Appl1
25	108.5	6.0	511	1	US-08-480-604A-20 Sequence 20, Appl1
26	108.5	6.0	511	2	US-08-405-496A-20 Sequence 20, Appl1
27	108.5	6.0	511	2	US-08-405-496A-20 Sequence 20, Appl1

28	108.5	6.0	608	1	US-08-480-604A-21 Sequence 21, Appl1
29	108.5	6.0	608	2	US-08-405-496A-21 Sequence 21, Appl1
30	108.5	6.0	609	1	US-08-480-604A-30 Sequence 30, Appl1
31	108.5	6.0	2366	1	US-08-480-604A-10 Sequence 10, Appl1
32	108.5	6.0	2366	2	US-08-405-496A-10 Sequence 10, Appl1
33	105	5.8	445	2	US-08-472-172-4 Sequence 4, Appl1
34	103.5	5.7	1794	6	5183745-6 Patent No. 5183745
35	102.5	5.7	1489	6	5183745-2 Patent No. 5183745
36	100.5	5.6	1577	2	US-08-793-824-2 Sequence 2, Appl1
37	100	5.5	1115	3	US-08-323-477-2 Sequence 2, Appl1
38	99.5	5.5	432	2	US-08-472-172-6 Sequence 6, Appl1
39	99.5	5.5	700	2	US-07-862-588B-2 Sequence 2, Appl1
40	99	5.5	459	2	US-08-472-172-2 Sequence 2, Appl1
41	98.5	5.4	617	1	US-08-361-920-29 Sequence 29, Appl1
42	98.5	5.4	617	1	US-08-479-939-29 Sequence 29, Appl1
43	98.5	5.4	617	1	US-08-483-432-29 Sequence 29, Appl1
44	94.5	5.2	607	4	US-08-537-361E-10 Sequence 10, Appl1
45	94.5	5.2	607	4	US-08-537-361E-10 Sequence 10, Appl1
					US-08-537-361E-10 Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-355-844-1
Sequence 1, Application US/08355844
Patent No. 5940307
GENERAL INFORMATION:
APPLICANT: Fischbary, Jorge
APPLICANT: Czegledy, Ferenc
APPLICANT: Iserovich, Pavel
APPLICANT: Li, Jun
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
TITLE OF INVENTION: STRUCTURE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/355,844
FILING DATE: 14-DEC-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Tang, Henry Y.S.
REGISTRATION NUMBER: 29,705
REFERENCE/DOCKET NUMBER: A29927-50/29910
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2586
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..340
OTHER INFORMATION: OMPF porin protein
US-08-355-844-1

Query Match 99.1%; Score 1792; DB 2; Length 340;
Best Local Similarity 99.4%; Pred. No. 2e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIVKDKNKVDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60
DB 1 AEIVKDKNKVDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60
QY WEYNQGNNSGADQOTGNKTRLAFLAGLYADVGSFDYGRNMGVYDLAGYTDMLPEFGG 120
DB 61 WEYNQGNNSGADQOTGNKTRLAFLAGLYADVGSFDYGRNMGVYDLAGYTDMLPEFGG 120
QY 121 DTAASDDEFVGVGAVATYRRNSNFFGLVDGLNFAVOYLGNKRDPTARRSNGDVGGSISY 180
DB 121 DTAASDDEFVGVGAVATYRRNSNFFGLVDGLNFAVOYLGNKRDPTARRSNGDVGGSISY 180
QY 181 EYEGFGIVGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNIYLAANYGETRNATPI 240
DB 181 EYEGFGIVGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNIYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYKSKAKDVEGIGVDLVNFEVGAT 300
DB 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYKSKAKDVEGIGVDLVNFEVGAT 300
QY 301 YFENKMSYVDYIIINOIDS DNKLGVSDDTYAVAGIYVQF 340
DB 301 YFENKMSYVDYIIINOIDS DNKLGVSDDTYAVAGIYVQF 340

RESULT 2
PCT-US95-16126-1
Sequence 1, Application PC/TUS9516126
GENERAL INFORMATION:
APPLICANT: Fischbarg, Jorge
APPLICANT: Czegledy, Ferenc
APPLICANT: Iserovich, Pavel
APPLICANT: Li, Jun
APPLICANT: Cheung, Min
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
TITLE OF INVENTION: STRUCTURE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/355,844
FILING DATE: 14-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Tang, Henry Y. S.
REGISTRATION NUMBER: 29,705
REFERENCE/DOCKET NUMBER: A29927-50/29910
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2586
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..340
OTHER INFORMATION: OmpF porin protein
PCT-US95-16126-1

Query Match 99.1%; Score 1792; DB 5; Length 340;
Best Local Similarity 99.4%; Pred. No. 2e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIVKDKNKVDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60
DB 1 AEIVKDKNKVDLYGKAVGLHYFSKNGENSYGGNGDMTYARLGFKGTQINSDLTGYGQ 60
QY 61 WEYNQGNNSGADQOTGNKTRLAFLAGLYADVGSFDYGRNMGVYDLAGYTDMLPEFGG 120
DB 61 WEYNQGNNSGADQOTGNKTRLAFLAGLYADVGSFDYGRNMGVYDLAGYTDMLPEFGG 120
QY 121 DTAASDDEFVGVGAVATYRRNSNFFGLVDGLNFAVOYLGNKRDPTARRSNGDVGGSISY 180
DB 121 DTAASDDEFVGVGAVATYRRNSNFFGLVDGLNFAVOYLGNKRDPTARRSNGDVGGSISY 180
QY 181 EYEGFGIVGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNIYLAANYGETRNATPI 240
DB 181 EYEGFGIVGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNIYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYKSKAKDVEGIGVDLVNFEVGAT 300
DB 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYKSKAKDVEGIGVDLVNFEVGAT 300
QY 301 YFENKMSYVDYIIINOIDS DNKLGVSDDTYAVAGIYVQF 340
DB 301 YFENKMSYVDYIIINOIDS DNKLGVSDDTYAVAGIYVQF 340

RESULT 3
US-09-041-889-30
Sequence 30, Application US/09041889
Patent No. 6033864
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
APPLICANT: Cohavy, Offer
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Peptide
US-09-041-889-30

Query Match 58.1%; Score 1051; DB 3; Length 367;
Best Local Similarity 58.9%; Pred. No. 2.9e-90;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;

QY 1 AEYKNDGNKVDLYKAVGLHYFSKNGENSYGNGDMTYARLGFGEQINSDLTGYGQ 60
DB 22 AEYKNDGNKVDLYKAVGLHYFSKNGENSYGNGDMTYARLGFGEQINSDLTGYGQ 76
QY 61 WEYNFGNNSGADAOCTGNKTRLAFLAGLKYADYSPDYGRNRYGVYDALGYDMLPEFG 120
DB 77 WEYQIGNSAEN---ENMSRTRAFAGLKRQDYGSPDYGRNRYGVYDVTWTDVLPPEFG 133
QY 121 DTAASDFVYGRVGYAATYRNSNPFGLVDGLNFAVOYLKNEBDT-----A 166
DB 134 DTAASDFVYGRVGYAATYRNSNPFGLVDGLNFAVOYLKNEBDT-----A 166
QY 167 RRSNGGVGGSISYEGFISGAYGADRTNIOE-AQPLNGKKAQWATGLKYDANNI 225
DB 194 LRONGGVGGSISYEGFISGAYGADRTNIOE-AQPLNGKKAQWATGLKYDANNI 225
QY 226 YLAANGTFRNATPTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADY-E 284
DB 254 YLAOYQTFYNNATRV-----GSLGMANKQNFVAOYQDFGLRPSIATYKSKADY-E 308
QY 285 GIDVDLVNFEVGYATYFNKNNSTYVDYIINOISDNKL-----GVGSDDTVA 340
DB 309 GYDDEDLKVDYGVATYFNKNNSTYVDYIINOISDNKL-----GVGSDDTVA 340

RESULT 4
US-09-041-889-29
Sequence 29, Application US/09041889
Patent No. 6033864

GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes thereof, using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-29

Query Match 56.9%; Score 1029; DB 3; Length 377;
Best Local Similarity 56.9%; Pred. No. 3.4e-88;
Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;

QY 1 AEYKNDGNKVDLYKAVGLHYFSKNGENSYGNGDMTYARLGFGEQINSDLTGYGQ 60
DB 22 AEYKNDGNKVDLYKAVGLHYFSKNGENSYGNGDMTYARLGFGEQINSDLTGYGQ 76
QY 61 WEYNFGNNSGADAOCTGNKTRLAFLAGLKYADYSPDYGRNRYGVYDALGYDMLPEFG 120
DB 77 WEYQIGNSAEN---ENMSRTRAFAGLKRQDYGSPDYGRNRYGVYDVTWTDVLPPEFG 133
QY 121 DTAASDFVYGRVGYAATYRNSNPFGLVDGLNFAVOYLKNEBDT-----A 166
DB 134 DTAASDFVYGRVGYAATYRNSNPFGLVDGLNFAVOYLKNEBDT-----A 166
QY 167 RRSNGGVGGSISYEGFISGAYGADRTNIOE-AQPLNGKKAQWATGLKYDANNI 225
DB 194 LRONGGVGGSISYEGFISGAYGADRTNIOE-AQPLNGKKAQWATGLKYDANNI 225
QY 226 YLAANGTFRNATPTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADY-E 284
DB 254 YLAOYQTFYNNATRV-----GSLGMANKQNFVAOYQDFGLRPSIATYKSKADY-E 308
QY 285 GIDVDLVNFEVGYATYFNKNNSTYVDYIINOISDNKL-----GVGSDDTVA 340
DB 309 GYDDEDLKVDYGVATYFNKNNSTYVDYIINOISDNKL-----GVGSDDTVA 340

RESULT 5
US-09-041-889-28
Sequence 28, Application US/09041889
Patent No. 6033864

GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes thereof, using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997

APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-28

Query Match 47.5%: Score 858.5: DB 3: Length 323;
Best Local Similarity 57.2%: Pred. No. 2.1e-72: Mismatches 72: Indels 25: Gaps 7;
Matches 174: Conservative 33: Mismatches 72: Indels 25: Gaps 7;

QY 1 AEYKDKGNKYDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60
DB 22 AEVYKDKGNKYDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 76
QY 61 WEYNQGNSEADQOTGKTRFLAFAGLKYADVSGFDYGRNRYVYDALGYDMLPEFGG 120
DB 77 WEYNQGNSEADQOTGKTRFLAFAGLKYADVSGFDYGRNRYVYDALGYDMLPEFGG 134
QY 121 DT-AVSDPEFVGRVGYATYRNSNFEGLVDGLNFAVQYLKNE-----RDTAR 167
DB 135 DSYTANDNFMTRANGVATYRNTDFEGLVNGLNFAVQYOGNNEGASNGDEGTNNGRD-VR 193
QY 168 RNSGDGVSISYEX-EGFGVGAAYGAADRNLQEAOPLGSKKAEDQWATGLKYANNY 226
DB 194 HENGQOMGLSTYYDLGMFSAGAAVYSSDRTNDQVNHMTAAGDKADAWTAGLKYANNY 253
QY 222 LAANGERNATPTINKFTNTSGFANKTQDVLVLAQYOPDFGLRPSIATYKSKADVECI 286
DB 254 LMTWSETRNMTPEGD---SDYAVANKTQNEFVTAQYDFGLRPAVSPLSKGRDLHAA 310
QY 287 GPVD 290
DB 311 GGAD 314

RESULT 6
US-08-096-182A-4
Sequence 4, Application US/08096182A
Patent No. 5439808
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huihan L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hironowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Porlin Proteins from Neisseria Meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,182A
FILING DATE: 23-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0060000
TELEPHONE: (202) 371-2500
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-182A-4

Query Match 8.7%: Score 157: DB 1: Length 342;
Best Local Similarity 24.7%: Pred. No. 6.1e-07: Mismatches 150: Indels 58: Gaps 16;
Matches 83: Conservative 45: Mismatches 150: Indels 58: Gaps 16;

QY 11 VLYGKAVGLHYFSKNGENSYGNGDMTY-----ARLFGKGTQINSDLTGYGQ 61
DB 3 VLYGKAVGLHYFSKNGENSYGNGDMTY-----ARLFGKGTQINSDLTGYGQ 62
QY 62 EYFQGNSEADQOTGKTRFLAFAGLKYADVSGFDYGR--NYGVYDALGYDMLPEFGG 119
DB 63 EYFQGNSEADQOTGKTRFLAFAGLKYADVSGFDYGR--NYGVYDALGYDMLPEFGG 113
QY 120 GDTAVSDPEFVGRVGYATYRNSNFEGLVDGLNFAVQYLKNE-----RDTAR 167
DB 135 DSYTANDNFMTRANGVATYRNTDFEGLVNGLNFAVQYOGNNEGASNGDEGTNNGRD-VR 193
QY 168 RNSGDGVSISYEX-EGFGVGAAYGAADRNLQEAOPLGSKKAEDQWATGLKYANNY 226
DB 194 HENGQOMGLSTYYDLGMFSAGAAVYSSDRTNDQVNHMTAAGDKADAWTAGLKYANNY 253
QY 222 LAANGERNATPTINKFTNTSGFANKTQDVLVLAQYOPDFGLRPSIATYKSKADVECI 286
DB 254 LMTWSETRNMTPEGD---SDYAVANKTQNEFVTAQYDFGLRPAVSPLSKGRDLHAA 310
QY 287 GPVD 290
DB 311 GGAD 314

RESULT 7
US-08-877-109-4
Sequence 4, Application US/08877109
Patent No. 5747287
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huihan L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hironowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Porlin Proteins from Neisseria meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., NW
CITY: Washington


```

1  APPLICANT: 12103 Indian Creek Court
2  APPLICANT: Beltsville, Maryland 20705
3  APPLICANT: United States of America
4  APPLICANT: Tal, Joseph Y.
5  APPLICANT: Qi, Huijin L.
6  APPLICANT: Liang, Shu-Mei
7  APPLICANT: Hironowski, Lucjan J. J.
8  APPLICANT: Pullen, Jeffrey K.
9  TITLE OF INVENTION: Method for the High Level
10 TITLE OF INVENTION: Expression,
11 TITLE OF INVENTION: Purification and Refolding of the Outer Membrane
12 TITLE OF INVENTION: Group B
13 TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
14 NUMBER OF SEQUENCES: 23
15 CORRESPONDENCE ADDRESS:
16 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
17 STREET: 1100 New York Ave., Suite 600
18 CITY: Washington
19 STATE: D.C.
20 COUNTRY: USA
21 ZIP: 20005-3934
22 COMPUTER READABLE FORM:
23 MEDIUM TYPE: Floppy disk
24 OPERATING SYSTEM: IBM PC compatible
25 SOFTWARE: Patentln Release #1.0, Version #1.25
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: PCT/US94/08327
28 FILING DATE: Herewith
29 CLASSIFICATION:
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/096,182
32 FILING DATE: 23 July 1993
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Esmond, Robert W.
35 REGISTRATION NUMBER: 32,893
36 REFERENCE/DOCKET NUMBER: 1438, 006PC00
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: (202) 371-2600
39 TELEFAX: (202) 371-2540
40 INFORMATION FOR SEQ ID NO: 4:
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 342 amino acids
43 TYPE: amino acid
44 TOPOLOGY: linear
45 MOLECULE TYPE: protein
46 PCT-US94-08327-4

```

[illegible]

```

QY      280 AKDVEGIGDVLVNY-----FEVGATYFNKNMSTYV 311
          | : | | | | | | | |
Db      283 KAKVNGVKD---ANQYDQYIVGADYDFSKRTSALV 315

```

```

10 RESULT
11 US-08-096-182A-6
12 : Sequence 6, Application US/08096182A
13 : Patent No. 5439808
14 : GENERAL INFORMATION:
15 : APPLICANT: Blake, Milan S.
16 : APPLICANT: Tai, Joseph Y.
17 : APPLICANT: Qi, Huijin L.
18 : APPLICANT: Liang, Shu-Mei
19 : APPLICANT: Hronowski, Lucjan J.J.
20 : APPLICANT: Pullen, Jeffrey K.
21 : TITLE OF INVENTION: Method for the High Level Expression,
22 : TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
23 : TITLE OF INVENTION: Purin Proteins from Neisseria Meningitidis
24 : NUMBER OF SEQUENCES: 23
25 : CORRESPONDENCE ADDRESS:
26 : ADDRESSEE: Sterne, Kessler, Goldstein & Fox
27 : STREET: 1100 New York Ave., Suite 600
28 : CITY: Washington
29 : STATE: D.C.
30 : COUNTRY: USA
31 : ZIP: 20005
32 : COMPUTER READABLE FORM:
33 : MEDIUM TYPE: Floppy disk
34 : COMPUTER: IBM PC compatible
35 : OPERATING SYSTEM: PC-DOS/MS-DOS
36 : SOFTWARE: Patentin Release #1.0, Version #1.25
37 : CURRENT APPLICATION DATA:
38 : APPLICATION NUMBER: US/08/096,182A
39 : FILING DATE: 23-Jul-1993
40 : CLASSIFICATION: 435
41 : ATTORNEY/AGENT INFORMATION:
42 : NAME: Esmond, Robert W.
43 : REGISTRATION NUMBER: 32,893
44 : REFERENCE/DOCKET NUMBER: 1438, 0060000
45 : TELECOMMUNICATION INFORMATION:
46 : TELEPHONE: (202) 371-2600
47 : TELEFAX: (202) 371-2540
48 : INFORMATION FOR SEQ ID NO: 6:
49 : SEQUENCE CHARACTERISTICS:
50 : LENGTH: 363 amino acids
51 : TYPE: amino acid
52 : TOPOLOGY: linear
53 : MOLECULE TYPE: protein
54 : US-08-096-182A-6

```

Query Match	8.7%	Score 157;	DB 1;	Length 363;
Best Local Similarity	24.7%	Pred. No. 6.7e-07;		
Matches	83;	Conservative	45;	Mismatches 150; Indels 58; Gaps 16;

OY	11	V	D	L	G	K	A	V	G	L	H	F	S	K	G	E	N	S	T	G	N	D	M	T	Y	-----	A	R	L	G	F	G	E	T	O	I	N	S	D	L	T	G	T	G	C	M									
Db	24	V	L	T	L	T	I	R	A	G	V	E	S	V	H	K	D	A	G	T	T	K	A	G	G	K	S	K	T	A	T	O	I	A	D	E	S	K	I	G	F	G	E	D	L	G	K	M	A	I	W	L			
OY	62	E	N	F	O	G	N	S	E	G	A	D	O	T	G	K	N	T	R	L	A	F	L	G	L	K	Y	A	D	V	S	E	F	D	Y	R	---	N	Y	G	V	U	Y	A	L	I	G	T	D	M	L	P	E	E	G
Db	84	E	---	Q	K	A	I	S	I	A	G	N	S	M	C	N	---	R	O	S	F	I	G	L	---	G	G	G	T	R	A	G	L	N	T	V	L	A	D	S	O	N	V	A	M	E	S	G							
OY	120	G	D	T	A	V	S	D	F	E	V	G	R	V	G	C	A	T	T	R	N	S	N	F	F	G	---	L	V	D	G	L	N	F	A	V	O	I	L	K	---	N	E	E	D	T	A	R	S	N	G	D	V		
Db	135	S	N	T	---	E	D	V	I	G	L	G	T	T	I	R	A	V	S	R	E	I	S	V	D	S	P	E	V	F	G	S	V	O	Y	A	R	D	N	A	N	D	V	K	K	K	H	T	R	S	S	R	E		
OY	175	G	S	I	S	E	Y	E	G	F	I	V	A	Y	G	A	---	A	D	R	---	T	N	L	O	E	A	O	P	L	N	G	K	K	A	E	O	M	A	T	G	L	K												
Db	193	S	H	A	L	K	T	E	N	G	F	E	O	I	Y	A	G	S	F	A	K	Y	A	D	L	M	T	D	E	R	V	A	N	T	A	N	H	P	---	K	D	Y	H	R	V	A	G	D							
OY	222	A	N	N	I	---	L	A	A	N	G	T	E	R	N	A	T	P	I	T	N	K	T	T	N	S	G	F	A	N	K	T	O	D	V	L	A	O	Y	O	R	D	E	G	L	R	S	I	A	T	T	K			

Db 250 ANDLYSVAGQYBAKN-----NEVGSTGKKHQQVATAAAYRFG-NVTPRVSTAHGF 303
QY 280 AKDVEGIGVDLVN-----FEVGATYYFNKNMSTYV 311
Db 304 KAKVNGVKD---ANYQYDOYIVGADYDFSKRTSALV 336

RESULT 11

US-08-877-109-6
; Sequence 6, Application US/0887109
; Patent No. 5747287
; GENERAL INFORMATION:
; APPLICANT: Blake, Milan S.
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Qi, Hulin L.
; APPLICANT: Liang, Shu-Mei
; APPLICANT: Hronowski, Lucjan J.J.
; APPLICANT: Pullen, Jeffrey K.
; TITLE OF INVENTION: Method for the High Level Expression,
; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/877,109
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,264
; FILING DATE: 28-APR-1995
; APPLICATION NUMBER: 08/096,182
; FILING DATE: 23-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0060001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-877-109-6

Query Match 8.7%; Score 157; DB 1; Length 363;
Best Local Similarity 24.7%; Pred. No. 6,7e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLYGAVGLHYFSKNGENSGNGDMY-----ARLFGKETQINSDLTGQW 61
Db 24 VTLYGIRKAGVEVSRVKAQGYKAGGSKSTATQIADFGSKIFKQEDLGNMKAIMWL 83
QY 62 EYNFGNNSGADAGQGNKRLAPAGLKAYDVGSFDYGR--NYGVVYDALGYDMLPEFG 119
Db 84 E--QKASTAGTSCWGN--ROSFILK---GGFSTVRAGNLTNVLKDSGDVNWAMESG 134
QY 120 GDTAVSDDPFVGVGVAVIYRNSNFG--LVQDLNFAVQYLK---NENDPTARNSGDGV 174

Db 135 SNT--EDVILGIGTIGRVESREISVRDSPVFAGFSSGVQVPRDNANDVDYKHKTRSSRE 192
QY 175 GGSISTEYEEFGFIVGAYGA-----ADR--TILQAOPLGKCKKAEONATGLKTD 221
Db 193 SYHAGLKYENAGFFGGVAGSPAKYADLNTDAERVAVNTANAHV---KDQVHRVAYGTD 249
QY 222 ANNLY--LAANYGETRNATPTINKFTNTSGFANKTODVLLVAQYOFDFGLRPSIAVTRSK 279
Db 250 ANDLYSVAGQYBAKN-----NEVGSTGKKHQQVATAAAYRFG-NVTPRVSTAHGF 303
QY 280 AKDVEGIGVDLVN-----FEVGATYYFNKNMSTYV 311
Db 304 KAKVNGVKD---ANYQYDOYIVGADYDFSKRTSALV 336

RESULT 12

US-08-798-760-6
; Sequence 6, Application US/08798760
; Patent No. 6013267
; GENERAL INFORMATION:
; APPLICANT: Blake, Milan S.
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Qi, Hulin L.
; APPLICANT: Liang, Shu-Mei
; APPLICANT: Hronowski, Lucjan J.J.
; APPLICANT: Pullen, Jeffrey K.
; TITLE OF INVENTION: Method for the High Level Expression,
; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,760
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0060002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-798-760-6

Query Match 8.7%; Score 157; DB 3; Length 363;
Best Local Similarity 24.7%; Pred. No. 6,7e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLYGAVGLHYFSKNGENSGNGDMY-----ARLFGKETQINSDLTGQW 61
Db 24 VTLYGIRKAGVEVSRVKAQGYKAGGSKSTATQIADFGSKIFKQEDLGNMKAIMWL 83
QY 62 EYNFGNNSGADAGQGNKRLAPAGLKAYDVGSFDYGR--NYGVVYDALGYDMLPEFG 119

Db 84 E---OKASIACTNSGWN--RQSFITGLK---GGFGTVFRAGNLTVLKDSGDNVNAMESG 134
120 GDTAYSDDFEYGRVGVATYRNSNFFG--LVYDGLNFAVOYLK---NERDARRSNGDGV 174
135 SNT--EDVLGIGTIGRVSRESRISVRYDSPVAFSGSVQVYPRNANDVDYKKTKSSRE 192
175 GGSISYEYEGFIVGAYGA-----ADR--TNLOEAPYLGNGKKAQMATGLKYD 221
193 SYHAGLKYEENAGFFGOYAGSPAKYADLNTDAERYAVNTANHPV---KDYQVHRVAVAGYD 249
222 ANNLY--LAANYGETRANTPTTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYTKSK 279
250 ANDLYSVAGQYEAKN-----NEVGSTKGKKHQTQVAAATPAATVAFG--NTPRVSYAHGF 303
280 AKDVEGIGDVLVNY---FEVGATYYFNKMMSTYV 311
304 KAKYNGVMD---ANYQYDQVTVGADYDPSKRTSALV 336

RESULT 13
PCT-US94-08327-6
Sequence 6, Application PC/TUS9408327
GENERAL INFORMATION:
APPLICANT: The Rockefeller University
APPLICANT: 1230 York Avenue
APPLICANT: New York, New York 10021
APPLICANT: United States of America
APPLICANT: 12103 Indian Creek Court
APPLICANT: Beltsville, Maryland 20705
APPLICANT: United States of America
APPLICANT: Tal, Joseph Y.
APPLICANT: Q1, HuiLin L.
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level
TITLE OF INVENTION: Purification and Refolding of the Outer Membrane
TITLE OF INVENTION: Group B
TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/08327
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,182
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438, 006PC00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US94-08327-6
Query Match 8.7%; Score 157; DB 5; Length 363;
Best Local Similarity 24.7%; Pred. No. 6,7e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;
11 VDLGKAVGLHYFSKNGENSYGNGDMTY-----ARLFGKETQIINSDLTGYGQW 61
24 VTLGYTIKAGVEVSRVVDAGTYKAGGSKSKATQIADFGSKIGFGQEDLGNKAIMQL 83
62 EYNFGNNSGEGADQOTNKPTLAFAGLKYADVGSPDGR--NYGVVDALGTYDMLPFG 119
84 E---OKASIACTNSGWN--RQSFITGLK---GGFGTVFRAGNLTVLKDSGDNVNAMESG 134
120 GDTAYSDDFEYGRVGVATYRNSNFFG--LVYDGLNFAVOYLK---NERDARRSNGDGV 174
135 SNT--EDVLGIGTIGRVSRESRISVRYDSPVAFSGSVQVYPRNANDVDYKKTKSSRE 192
175 GGSISYEYEGFIVGAYGA-----ADR--TNLOEAPYLGNGKKAQMATGLKYD 221
193 SYHAGLKYEENAGFFGOYAGSPAKYADLNTDAERYAVNTANHPV---KDYQVHRVAVAGYD 249
222 ANNLY--LAANYGETRANTPTTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYTKSK 279
250 ANDLYSVAGQYEAKN-----NEVGSTKGKKHQTQVAAATPAATVAFG--NTPRVSYAHGF 303
280 AKDVEGIGDVLVNY---FEVGATYYFNKMMSTYV 311
304 KAKYNGVMD---ANYQYDQVTVGADYDPSKRTSALV 336

RESULT 14
US-08-096-182A-2
Sequence 2, Application US/08096182A
Patent No. 5439808
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tal, Joseph Y.
APPLICANT: Q1, HuiLin L.
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
TITLE OF INVENTION: Purification and Refolding of the Outer Membrane
TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,182A
FILING DATE: 23-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438, 0060000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-182A-2

Query Match 8.2%; Score 147.5; DB 1; Length 309;
Best Local Similarity 25.1%; Pred. No. 4.1e-06;
Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;

35 NGDMTY-----ARLGFKEGTQINSDLTGQWMEYFPGNNSGADAGTGNKTRL 83
19 NGQVTEVTATGIVDLSKIGFGQEDLGNLKAIVQVE---QKASIACTDSGMN--RQ 73
QY 84 AFAGLKYADVGSFDYGRNYGVYDALGYTDMLEPFGD---TAVSDDFVGRVGG---- 135
DB 74 SFTGLK-GGFGKLRVGRLSVLKDT-----GDINPMDSKSDYLGVNKIAPEAR 121
QY 136 --VATYRNSNFEGLVDGLNFAVOYLGNKNERDARNSNGDVGGSISYEYEGFIVAGYA 193
DB 122 LISVRIDSPFEAGLSGSVOYAL-----NDNAGRHNSESYHAGFVYKKGCFV--QYGG 172
QY 194 ADRTNLOEAQPLGNCK-KAEQWATGLKYDANNIYLAANGETRNPATPINKFTNTSGFAN 252
DB 173 AKRHHQVOEGELIEKYQIHRIVSG--YNDALY-ASVAVQODA-----KLTDAASNHN 224
QY 253 KTDVLLVAQYQDFGLRPSIATYTSKAKDVEG-IGDVLVNYFE--VGATYFENKNS 308
DB 225 SQTEVAATLAVRFG-NVTPRVSY---AHGFKGLVDADICNEIDYDVVGAEDVFSKRTS 279
QY 309 TYV 311
DB 280 ALV 282

RESULT 15

US-08-877-109-2
Sequence 2, Application US/08877109

GENERAL INFORMATION:

APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Qi, Huijin L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hronowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/877.109
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/431.264
FILING DATE: 28-APR-1995
APPLICATION NUMBER: 08/096.182
FILING DATE: 23-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Edmond, Robert W.

REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0060001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-877-109-2

Query Match 8.2%; Score 147.5; DB 1; Length 309;
Best Local Similarity 25.1%; Pred. No. 4.1e-06;
Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;

35 NGDMTY-----ARLGFKEGTQINSDLTGQWMEYFPGNNSGADAGTGNKTRL 83
19 NGQVTEVTATGIVDLSKIGFGQEDLGNLKAIVQVE---QKASIACTDSGMN--RQ 73
QY 84 AFAGLKYADVGSFDYGRNYGVYDALGYTDMLEPFGD---TAVSDDFVGRVGG---- 135
DB 74 SFTGLK-GGFGKLRVGRLSVLKDT-----GDINPMDSKSDYLGVNKIAPEAR 121
QY 136 --VATYRNSNFEGLVDGLNFAVOYLGNKNERDARNSNGDVGGSISYEYEGFIVAGYA 193
DB 122 LISVRIDSPFEAGLSGSVOYAL-----NDNAGRHNSESYHAGFVYKKGCFV--QYGG 172
QY 194 ADRTNLOEAQPLGNCK-KAEQWATGLKYDANNIYLAANGETRNPATPINKFTNTSGFAN 252
DB 173 AKRHHQVOEGELIEKYQIHRIVSG--YNDALY-ASVAVQODA-----KLTDAASNHN 224
QY 253 KTDVLLVAQYQDFGLRPSIATYTSKAKDVEG-IGDVLVNYFE--VGATYFENKNS 308
DB 225 SQTEVAATLAVRFG-NVTPRVSY---AHGFKGLVDADICNEIDYDVVGAEDVFSKRTS 279
QY 309 TYV 311
DB 280 ALV 282

Search completed: July 3, 2001, 14:56:46
Job time: 536 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:59:45 ; Search time 63.58 Seconds
(without alignments)
408.549 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1809

Sequence: 1 AEINRKGDKVLDYKAVGL.....NKLGVGSDDTVAAGIVQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	1	MMECF
2	1789	98.9	362	2	B85619
3	1113	61.5	351	2	B85513
4	1110	61.4	351	1	MMECFE
5	1070	59.1	351	2	S25520
6	1058	58.5	350	2	S25525
7	1051	58.1	365	1	MMBP2
8	1051	58.1	367	1	MMECP2
9	1044	57.7	350	2	S36606
10	1042	57.6	365	1	MMECP2
11	1035	57.2	377	2	D64888
12	1033	57.1	367	2	B85861
13	1030	56.9	363	2	S51104
14	1029.5	56.6	363	2	S43159
15	1023.5	56.6	378	1	MMEBP2
16	1000.5	55.3	378	1	A59139
17	998.5	55.2	366	2	E85729
18	976.5	54.0	394	2	S34263
19	885	48.9	369	2	S70847
20	770	42.6	315	2	T46993
21	635.5	35.1	255	2	D85740
22	573.5	31.7	382	2	G84971
23	539.5	29.8	224	2	F85814
24	486	26.9	171	2	H64960
25	415	22.9	191	2	H64960
26	372.5	20.6	123	2	C85740
27	351	19.4	134	2	B64961
28	236.5	13.1	350	2	A82399
29	233.5	12.9	96	2	C64900

30	224	12.4	331	2	E82389	probable outer mem
31	197.5	10.9	330	2	PC4425	lectin-like adhesi
32	164.5	9.1	97	2	A57983	outer membrane por
33	157	8.7	360	2	S17456	outer membrane por
34	157	8.7	361	2	JC6314	major porin protei
35	155.5	8.6	50	2	T46990	hypothetical prote
36	153.5	8.5	312	2	PH0224	class-3 porin prot
37	153.5	8.5	371	2	S68069	major outer membra
38	151.5	8.4	371	2	S68072	major outer membra
39	150.5	8.3	313	2	JH0254	class-3 porin prot
40	150.5	8.3	331	2	S21407	major outer membra
41	150.5	8.3	390	2	I39570	probable porin (s1
42	149	8.2	323	2	S25514	outer membrane pro
43	147.5	8.2	295	2	S25507	outer membrane pro
44	144.5	8.0	295	2	S25506	outer membrane pro
45	139.5	7.7	359	2	D64050	major outer membra

ALIGNMENTS

RESULT 1

MMECF

outer membrane porin ompF precursor - Escherichia coli

N:Alternate names: outer membrane protein 1a; outer membrane protein b

C:Species: Escherichia coli

C>Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 16-Jul-1999

C/Accession: A93449; A91301; A90314; A25029; H64832; A03431

R:Inokuchi, K.; Mutoh, N.; Matsuyama, S.; Mizushima, S.

Nucleic Acids Res. 10, 6957-6968, 1982

A:Title: Primary structure of the ompF gene that codes for a major outer membrane pro

A:Reference number: A93449; PMID:83050452

A:Accession: A93449

A:Residues: 1-362 <INOS>

A:Cross-references: GB:J01655; GB:M10311; GB:M10312; NID:g147009; PIDN:AAA24244.1; PI

A:Experimental source: strain K12

R:Muton, N.; Inokuchi, K.; Mizushima, S.

FEBS Lett. 137, 171-174, 1982

A:Title: Amino acid sequence of the signal peptide of OmpF, a major outer membrane pr

A:Reference number: A91301; PMID:82139379

A:Accession: A91301

A:Molecule type: DNA

A:Residues: 1-37 <MOT>

R:Chen, R.; Kramer, C.; Schmidmayr, W.; Chen-Schmeisser, U.; Henning, U.

Biochem. J. 203, 33-43, 1982

A:Title: Primary structure of major outer-membrane protein I (ompF protein, porin) of

A:Reference number: A90314; PMID:82256494

A:Accession: A90314

A:Molecule type: protein

A:Residues: 23-87, 'E', 89-138, 'G', 140-283, 'L', 285-362 <CHB>

R:Nogami, T.; Mizuno, T.; Mizushima, S.

J. Bacteriol. 164, 797-801, 1985

A:Title: Construction of a series of ompF-ompC chimeric genes by in vivo homologous r

A:Reference number: A91809; PMID:86033642

A:Accession: A25029

A:Molecule type: DNA

A:Residues: 33-63 <NOG>

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617

A:Accession: H64832

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-362 <BAT>

A:Cross-references: GB:AE000195; GB:U00096; NID:g1787156; PIDN:AACT4015.1; PID:g17871

A:Experimental source: strain K-12, substrain MG1655

C/Genetics:

A:Gene: ompF; tolF; cmfB; coa; cry

A:Map position: 21 min

C:Complex: homotrimer

C:Function: <POR>
 A:Description: forms aqueous channels that facilitate diffusion of small hydrophilic mole
 C:Function: <REC>
 A:Description: receptor for bacteriophage T2
 C:Superfamily: outer membrane protein phoe
 C:Keywords: membrane protein; porin; trimmer
 F:1-22/Domain: signal sequence #status predicted <Sig>
 F:23-362/Product: outer membrane porin ompF #status experimental <Mat>

Query Match 99.8%; Score 1805; DB 1; Length 362;
 Best Local Similarity 100.0%; Pred. No. 3e-123;

Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AEIYNKDKGNKVDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTYGQ 60
 |||||||
 Db 23 AEIYNKDKGNKVDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTYGQ 82
 |||||||
 QY 61 WEYNFOGNNSEGADAOGTGNKTRLAFLAKYADVGSFDYGRNYGVVYDALGYTDLMPFEGG 120
 |||||||
 Db 83 WEYNFOGNNSEGADAOGTGNKTRLAFLAKYADVGSFDYGRNYGVVYDALGYTDLMPFEGG 142
 |||||||
 QY 121 DTAASDPEFGRGVAVATYRNSNFEGLVDGLNFAVOYLGNERNPTARNSGDVGGSISY 180
 |||||||
 Db 143 DTAASDPEFGRGVAVATYRNSNFEGLVDGLNFAVOYLGNERNPTARNSGDVGGSISY 202
 |||||||
 QY 181 EYEGFGIVAGYGAADRTNLOEAQPLGNGKKAQMAWATGLKYDANNIYLAANYGETRNPPI 240
 |||||||
 Db 203 EYEGFGIVAGYGAADRTNLOEAQPLGNGKKAQMAWATGLKYDANNIYLAANYGETRNPPI 262
 |||||||
 QY 241 TNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADVEGIGDVLVNFYEVGAT 300
 |||||||
 Db 263 TNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADVEGIGDVLVNFYEVGAT 322
 |||||||
 QY 301 YFFNKNMSTYVDYIIINOISDNKLGVSDDPTVAVGITYQF 340
 |||||||
 Db 323 YFFNKNMSTYVDYIIINOISDNKLGVSDDPTVAVGITYQF 362
 |||||||

RESULT 2
 B85619
 outer membrane protein Ia (ta,b,F) [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 31-Mar-2001

C:Accession: B85619
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca,
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: B85619
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-362 <STO>

A:Cross-references: GB:A8005174; NID:912514104; PIDN:AAG55414.1; GSPDB:GN00145; UMGF:Z12

A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ompF

C:Superfamily: outer membrane protein phoe

Query Match 98.9%; Score 1789; DB 2; Length 362;
 Best Local Similarity 99.4%; Pred. No. 4.3e-122;

Matches 338; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEIYNKDKGNKVDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTYGQ 60
 |||||||
 Db 23 AEIYNKDKGNKVDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTYGQ 82
 |||||||
 QY 61 WEYNFOGNNSEGADAOGTGNKTRLAFLAKYADVGSFDYGRNYGVVYDALGYTDLMPFEGG 120
 |||||||
 Db 83 WEYNFOGNNSEGADAOGTGNKTRLAFLAKYADVGSFDYGRNYGVVYDALGYTDLMPFEGG 142
 |||||||

QY 121 DTAASDPEFGRGVAVATYRNSNFEGLVDGLNFAVOYLGNERNPTARNSGDVGGSISY 180
 |||||||
 Db 143 DTAASDPEFGRGVAVATYRNSNFEGLVDGLNFAVOYLGNERNPTARNSGDVGGSISY 202
 |||||||
 QY 181 EYEGFGIVAGYGAADRTNLOEAQPLGNGKKAQMAWATGLKYDANNIYLAANYGETRNPPI 240
 |||||||
 Db 203 EYEGFGIVAGYGAADRTNLOEAQPLGNGKKAQMAWATGLKYDANNIYLAANYGETRNPPI 262
 |||||||
 QY 241 TNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADVEGIGDVLVNFYEVGAT 300
 |||||||
 Db 263 TNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADVEGIGDVLVNFYEVGAT 322
 |||||||
 QY 301 YFFNKNMSTYVDYIIINOISDNKLGVSDDPTVAVGITYQF 340
 |||||||
 Db 323 YFFNKNMSTYVDYIIINOISDNKLGVSDDPTVAVGITYQF 362
 |||||||

RESULT 3

B85613
 outer membrane pore protein E (E.Ic.Nmpab) [imported] - Escherichia coli (strain O157

C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 31-Mar-2001

C:Accession: B85613
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apoda
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: B85613
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-351 <STO>

A:Cross-references: GB:A8005174; NID:912513015; PIDN:AAG54566.1; GSPDB:GN00145; UMGF:

A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: phoe

C:Superfamily: outer membrane protein phoe

Query Match 61.5%; Score 1113; DB 2; Length 351;
 Best Local Similarity 62.1%; Pred. No. 2.9e-73;

Matches 213; Conservative 38; Mismatches 76; Indels 16; Gaps 5;

QY 1 AEIYNKDKGNKVDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTYGQ 60
 |||||||
 Db 22 AEIYNKDKGNKVDLYKAVGLHYFSKNGENSYGGNGDMTYARLGFKEGTQINSDLTYGQ 76
 |||||||
 QY 61 WEYNFOGNNSEGADAOGTGNKTRLAFLAKYADVGSFDYGRNYGVVYDALGYTDLMPFEGG 120
 |||||||
 Db 77 WEYNFOGNNSEGADAOGTGNKTRLAFLAKYADVGSFDYGRNYGVVYDALGYTDLMPFEGG 134
 |||||||
 QY 121 D-TAASDPEFGRGVAVATYRNSNFEGLVDGLNFAVOYLGNERNPTARNSGDVGGSIS 179
 |||||||
 Db 135 DSAQTDNFMKRRASGLATYRTDFEGLVIDLNTLQYOGKNNRNDVAKONGDGFISLT 194
 |||||||
 QY 180 YEYES--FGIYAGYGAADRTNLOEAQPLGNGKKAQMAWATGLKYDANNIYLAANYGETRNPPI 237
 |||||||
 Db 195 YDFGSDFAISGATYNSDRTNLOEAQPLGNGKKAQMAWATGLKYDANNIYLAANYGETRNPPI 254
 |||||||
 QY 238 TPITNKTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADVEGIGDVLVNFYEVGAT 297
 |||||||
 Db 255 TPI-----TGGFANKTQDVEAVAYQDFGLRPSIATYKSKADVEGIGDVLVNFYEVGAT 308
 |||||||
 QY 298 GATYFFNKNMSTYVDYIIINOISDNKLGVSDDPTVAVGITYQF 340
 |||||||
 Db 309 GATYFFNKNMSTYVDYIIINOISDNKLGVSDDPTVAVGITYQF 351
 |||||||

RESULT 4

NMECE
 outer membrane porin phoe precursor - Escherichia coli

C:Species: Escherichia coli
 C:Date: 13-Jun-1983 #sequence-revision 13-Jun-1983 #text-change 21-Jul-2000


```

Db 77 WEYQIOGNSEN---ENNSTRAFAAGLKFQDVGSPDYGNKYVDVYTSWTVLPEFGG 133
Oy 121 DPAVSDDFEVGRVGVATYRNSNPFGLVDGLNFAVOYLKNERDT-----A 166
Db 134 DTVGSDFNPMQORNGNATYRNTDFGLVDGLNFAVOYOGKNGSGEGFTSGVTNNGRDA 193
Oy 167 RNSNGDVGGSISYEYEGGIGVATGADRTNIOE-AQPLGSKKAQWATGKTYDANNI 225
Db 194 LRONGDVGGSITYDEYEGGIGCAISSKRTDAONTAAIYGNDRRAETGYGGLKYDANNI 253
Oy 226 YLAANYGGERNATPITNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATKSKARDV-E 284
Db 254 YLAQYTTQYNNARV-----GSLGMANKAQNFAVAOYQDFGLRPSLATLQSKGNLGR 308
Oy 285 GIGDVLVYFEVATYFNKNSMTYVDYIINOISDNKL-----GVGSDPTVAVGVIYQF 340
Db 309 GYDDEDILKTVYDGVATYFNKNSMTYVDYKINLDD-DNQFTRRAGINTDNIVALGLVYQF 367

RESULT 9
outer membrane porin phoe precursor - Salmonella typhi
C:Species: Salmonella typhi
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C:Accession: S36606
R:Torres, A.; Puente, J.L.; Calva, E.
Submitted to the EMBL Data Library, August 1993
A:Reference number: S36606
A:Accession: S36606
A:Molecule type: DNA
A:Residues: 1-350 <TOR>
A:Cross-references: EMBL:X74595; NID:g397159; PIDN:CA52672.1; PID:g397160
A:Experimental source: Isolate IMS-1
A:Note: this is one of the proteins induced when the E. coli cells are grown under phospho
d compounds, and some other negatively charged solutes
C:Genetics:
A:Gene: phoe
C:Complex:
A:Description: allows passive diffusion of small hydrophilic molecules with molecular we
A:Note: Induced under phosphate limitation
C:Superfamily: outer membrane protein phoe
C:Keywords: membrane protein; porin
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-350/Product: outer membrane porin phoe #status predicted <OPP>

Query Match 57.7%; Score 1044; DB 2; Length 350;
Best Local Similarity 57.4%; Pred. No. 2.8e-68;
Matches 197; Conservative 47; Mismatches 83; Indels 16; Gaps 5;

Oy 1 AEIYNKDNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFGEINOINSLDLYGQ 60
Db 21 AEIYNKNGKLDYGVKAMHYSDYDSK-----DGDQSYVRGFGKGTIOINQLTGYGR 75
Oy 61 WEYNFOGNNSEGAQDAQGNKTRLAFAGLKAYADVSPDYGRNRYGVVYDALCYTDLPEFGG 120
Db 76 WEYEGAGNKAESDSQ--OKTRLAFAGLKLDIGSFQGNLICALYDEAMTQMPPEFGG 133
Oy 121 D-TAYSDDFVGRVGVATYRNSNPFGLVDGLNFAVOYLKNERDRAARNSNGVGSIS 179
Db 134 DSSAQDQNTMTKRAASGLATYRNTDFGLVDGLTLOLYOGKNDRDVKKONGGSGFTGSVS 193
Oy 180 YEYEG--FCIGVAGYGAADRTNLQEAQPLGSKKAQWATGKTYDANNIYLAANYGETRNA 237
Db 194 YDGGSDPAVSGATYTSDRTRRQNLQRTGTDAAEGHATCVKTDADITYATYSETRNA 253
Oy 238 TPTTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYTSKAKADVEGIGDVLVNYFEV 297
Db 254 TPV-----SGFRANKTQNEAVIYOQDFGLRPSLGYVLSKSKDIEGVGSEDLVNYIDV 307
Oy 298 GATYFNKNSMTYVDYIINOISDNKLGVGSDPTVAVGVIYQF 340

```

```

Db 308 GATYFNKNSAFYDKIKINQDSDNTLGIINDDIYALGLYQF 350

RESULT 10
outer membrane porin mmpc precursor - Escherichia coli cryptic lambdaoid prophage DLP1
MECNC
C:Species: Escherichia coli
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A25647; S66594; G64787
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaltman, C.A.
J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the lc and mmpc outer membrane porin protein genes of lambdaoid
A:Reference number: A25647; MUID:86304457
A:Contents: mutant strain CS384
A:Accession: A25647
A:Molecule type: DNA
A:Residues: 1-365 <BLA>
R:Mandt, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.
J. Mol. Biol. 257, 561-573, 1996
A:Title: Holliday junction resolvases encoded by homologous rusa genes in Escherichia
A:Reference number: S66579; MUID:96196428
A:Accession: S66594
A:Molecule type: DNA
A:Residues: 347-365 <MAH>
A:Cross-references: EMBL:X92587; NID:g1051136; PIDN:CA63325.1; PID:g1051145
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64787
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'MNIRAVTSPFNNSKKGLT', 1-325, 'N', 327-347, 'EGANKSLI' <BLAT>
A:Cross-references: GB:AE000160; GB:U00096; NID:g1786751; PIDN:AC03654.1; PID:g17867
A:Experimental source: strain K-12, substrain MG1655
A:Comment: In wild-type strains of E. coli K-12, the mmpc open reading frame is inter
ain CS348, the 155 element has been deleted and mmpc is expressed.
C:Genetics:
A:Gene: mmpc
A:Map position: 12 min
A:Genome: cryptic lambdaoid prophage DLP12
C:Superfamily: outer membrane protein phoe
C:Keywords: membrane protein; porin; trimer
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin mmpc #status predicted <MAT>

Query Match 57.6%; Score 1042; DB 1; Length 365;
Best Local Similarity 59.5%; Pred. No. 4.2e-68;
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;

Oy 1 AEIYNKDNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFGEINOINSLDLYGQ 60
Db 24 AEIYNKDSNKLDTLGKNAKHFSSNDAD-----DGYTTRARLGFGEIYNOQLTGFQ 78
Oy 61 WEYNFOGNNSEGAQDAQGNKTRLAFAGLKAYADVSPDYGRNRYGVVYDALCYTDLPEFGG 120
Db 79 WEYEPGNRAE-SGSSKQDTRLAFAGLKFGDYSIDYGNRYGVAVDIGAMTVLPEFGG 137
Oy 121 DT-AYSDDFVGRVGVATYRNSNPFGLVDGLNFAVOYLKNER--DTARRNSNGDVG 176
Db 138 DFTWQTDVFMTORATGATYRNTDFGLVDGLNFAVOYQKKNRSDPDNTTEGNGJGFGF 197
Oy 177 SISEYEGGIGVATGADRTNIOE-----AQPLGSKKAQWATGKTYDANNIYLAAN 230
Db 198 SATYEEGFGIGATYAKSDRTDQVNAAGVLPYFASGKRAEWAAGLKYDANNIYLAAT 257
Oy 231 YGETRNATPITNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYTSKAKADVEGIGD 290
Db 258 YSETOKNTYADHFV-----ANKAQNFEAVIYOQDFGLRPSVAIYLSKSKDLGVWGDD 312

```


Qy	61	MEYNQNNNSGADQOTENKRLTAFLAGLKYADVSFEQYGRNVGVYDMLAGTDMPEEGG	120
Dp	77	MEYNQANNTESSSQOAM--TRLPAFLKFGDAGSEFDGRNIGVYDWTSMTDLPPEGG	134
Qy	121	DTAASDDEFVGRGVGATYRNSNFEGLVDGLNFAVOYLGNK-----ERDT-----ARRSN	170
Dp	135	DTYGSDFNLFQSANCVATYRNSNDFEGLVDGLNFALOYQGNKGSVSGEGATNNGRCALKON	194
Qy	171	GDGVGGSISYE-YEEFGYGVAGAGADRTNLQEAQPLGNGKKAECMAETLTKTDANNIYLA	229
Dp	195	GDGFQTSYTYDIFDISAGFPAFANSKRITDDONQOLLGEGDHAEYTGGLKTDANNIYLAT	254
Qy	230	NYGETRNATPIITNKETNTNCSGFANKQTODVLLAAYQFDEGLRPSIAYTSKAKADYEGID	289
Dp	255	QYTQTYNATR-----AGSLGFANKQNEVAAYQYFDEGLRPSVAYLQSKGKDLNGYEDQ	309
Qy	290	DLVNTFEVGGATYTYFNKKNMSTYVDYLIINOISDND---KGVGSDPYVANGIYYQF	340
Dp	310	DILKAYVDGATYTYFNKKNMSTYVDYKINLLDNDSFTRSGAISTDDVVALGLVYQF	363

RESULT 14
S43159
outer membrane porin ompF precursor - *Salmonella typhimurium*
C:Species: *Salmonella typhimurium*
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C:Accession: S43159
R:Venegas, A.; Gomez, I.; Bruce, E.; Martinez, M.
submitted to the EMBL Data Library, March 1994
A:Description: PCR amplification and cloning of the *Salmonella typhimurium* ompF porin gene
A:Reference number: S43159
A:Accession: S43159
A:Molecule type: DNA
A:Residues: 1-363 <VEN>
C:Cross-references: EMBL:Z31594; NTD:9468740; PIDD:CA83471.1; PTD:9468741
C:Genetics:
A:Gene: ompF
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-32/Domain: signal sequence #status predicted <SIG>
F:23-363/Product: outer membrane porin ompF #status predicted <NAT>

[illegible]

RESULT 15

MMEBPC
 Outer membrane porin ompC precursor - Salmonella typhi
 C:Species: Salmonella typhi
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
 C:Accession: J00119; S01248; S10122
 R:Punte, J.L.; Alvarez-Scherer, V.; Gosset, G.; Calva, E.
 Gene 83, 197-206, 1989
 A>Title: Comparative analysis of the Salmonella typhi and Escherichia coli ompC genes
 A:Reference number: J00119; MUID:90060831
 A:Accession: J00119
 A:Molecule type: DNA
 A:Residues: 1-378 <PUE>
 A:Cross-references: GB:M1424; NID:g154207; PIDN:AAA27169.1; PID:g154208
 R:Venegas, A.; Gomez, I.; Zaror, I.; Yudelevich, A.
 Nucleic Acids Res. 16, 7721, 1988
 A>Title: The nucleotide sequence of the Salmonella typhi ompC porin gene.
 A:Reference number: S01248; MUID:88319959
 A:Accession: S01248
 A:Molecule type: DNA
 A:Residues: 1-5,'Q','7-361,363-378 <VEN1>
 A:Cross-references: EMBL:X07835; NID:g47796; PIDN:CAA30688.1; PID:g47797
 A:Experimental source: strain Ty2
 R:Venegas, A.
 submitted to the EMBL Data Library, May 1988
 A:Reference number: S10122
 A:Accession: S10122
 A:Molecule type: DNA
 A:Residues: 1-361,363-378 <VEN2>
 A:Cross-references: EMBL:X07835; NID:g47796; PIDN:CAA30688.1; PID:g47797
 C:Genetics:
 A:Gene: ompC
 A:Map position: 45 min
 C:Superfamily: Outer membrane protein phoe
 C:Keywords: membrane protein, porin, trimer
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:21-378/Product: outer membrane porin ompC #status predicted <ANT>

Query Match	56.6%	Score 1023.5	DB 1	Length 378
Best Local Similarity	57.38	Pred. No. 9,4e-67		
Matches 209	Conservative	44	Mismatches 79	Indels 33
				Gaps 10
QY 1 AEIYKKNDDKKNVLYLAKAGVLAHYFSKNGCNGENSXGCGNDMTYALGKGEQTOINSDLTGCGQ 60				
Db 22 AEIYKKNDDKKNKLDLPEKIVSLGHYFS-----DKGSGDGDQYVMIIGKGEQVVDQLTIGVQ 76				
QY 61 WEYNFOGNNBEGADADOTGKTKRLAFAGLKYADVSGEDYGRNTGVYYDALGYDMLEPFGG 120				
Db 77 WEYTOJOGNQTGESSN--DSMTVFAFAGLFLADAGSGEDYGRNVTYDVTSMWDVLDPEFGG 133				
QY 121 DPAVSDDFPVGVGVGVAATYARNSPFGLVDYGLNFAVQYLGN-----ERDTRK--RSGND 172				
Db 134 DTYGADNFPQQRNGNATYATRTDTPFGLDVGLDFALQYQKNSVSGEENTNGSLNMGND 193				
QY 173 GVGGSISLEY-BGFIIVGAY---GAADRTNLQEAOPLGNGKRAEQMATGLKYDANNIYL 227				
Db 194 GYGGSILTYAIGGFSVGAITTSKRTADONNANRLYNGDRAVYTGGLKYDANNIYL 253				
QY 228 AANVGETRRAATPI--NNKPTNTS--GFANKQDVLVLYAOYQDFGLRPAITATKSKADY 283				
Db 254 AAOYQOTYVATPFGTSNGSNSTSYGFAPKAQNFEEVVAOYQDFGLRPSVALVQSKGKDI 313				
QY 284 E-----GIDGVLDVNVFEVGAATYFNKNKNSITVDYLIINOIDSDN--KLGVASDQTFVAG 335				
Db 314 SNGVGASVGDQDQIVAYYVDGATYIYFNKNKNSITVDYKINILDKNDNPTROAGINTVDDIVALG 373				
QY 336 IYQOF 340				
Db 374 IYQOF 378				

Search completed: July 3, 2001, 14:59:46

Job time: 371 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:50 ; Search time 37.35 Seconds
(without alignments)
312.748 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1809

Sequence: 1 AEIYNKGNVDLXGRAVGL.....NKLGVSDPIYVAVGIYVQFA 341

Scoring table: BIOSUM62
Gap 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	1	OMP_F_ECOLI
2	1110	61.4	351	1	P02931 escherichia
3	1072.5	59.3	350	1	PHOE_ENTCL
4	1072	59.3	374	1	OMP_F_SERMA
5	1070	55.1	351	1	PHOE_CITFR
6	1058	58.5	350	1	PHOE_SALTY
7	1051	58.1	367	1	OMP_C_ECOLI
8	1050	58.0	351	1	PHOE_KLEPN
9	1046	57.8	365	1	P07238 bacterioph
10	1044	57.7	350	1	PHOE_SALTY
11	1042	57.6	365	1	NMPC_ECOLI
12	1041	57.5	349	1	PHOE_KLEOX
13	1035	57.2	377	1	OMP_N_ECOLI
14	1031.5	57.0	363	1	OMP_F_SALTY
15	1030	56.9	363	1	OMP_C_KLEPN
16	1029.5	56.9	363	1	OMP_F_SALTY
17	1027	56.8	383	1	OMP_F_SALTY
18	1023.5	55.6	378	1	OMP_C_SALTY
19	1000.5	55.3	378	1	OMP_C_SALTY
20	999	55.2	397	1	YEDS_ECOLI
21	976.5	54.0	394	1	OMSL_SALTY
22	902	49.9	342	1	OMP_C_RAHMO
23	894.5	49.4	376	1	OMP_C_SERMA
24	885	48.9	369	1	OMP_F_XENNE
25	825.5	45.6	282	1	OMP_F_SALTY
26	573.5	31.7	382	1	P0RL_BUCAT
27	236.5	13.1	341	1	OMP_U_VIBCH
28	218	12.1	341	1	OMP_L_PHOPR
29	170.5	9.4	353	1	OML2_HAEIN
30	150.5	8.4	371	1	OML2_HAEIN
31	141.5	7.8	360	1	OML2_HAEIN
32	141.5	7.7	359	1	OML2_HAEIN
33	139.5	7.7	359	1	OML2_HAEIN

34	139	7.7	361	1	OML2_HAEIN	048217 haemophilus
35	138.5	7.7	348	1	OML1_NEIGO	P18195 neisseria g
36	138.5	7.7	361	1	OML2_HAEIN	P20148 haemophilus
37	135.5	7.5	369	1	OML2_HAEIN	048218 haemophilus
38	133	7.4	1829	1	FRPC_NEIMB	0919V5 neisseria m
39	131.5	7.3	348	1	OML2_NEIGO	P30688 neisseria m
40	131	7.2	331	1	OML2_NEIMB	P55127 neisseria m
41	131	7.2	1829	1	FRPC_NEIMC	P55127 neisseria m
42	130	7.2	367	1	OML2_HAEIN	048219 haemophilus
43	129.5	7.2	363	1	OML2_HAEIN	048220 haemophilus
44	129	7.1	331	1	OML1_NEIMB	P30687 neisseria m
45	129	7.1	331	1	OML3_NEIMB	P30689 neisseria m

ALIGNMENTS

RESULT	ID	OMP_F_ECOLI	STANDARD	PRT	362 AA.
AC	P02931				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	OUTER MEMBRANE PROTEIN F PRECURSOR (PORIN OMPF) (OUTER MEMBRANE PROTEIN 1A, IA, OR B).				
GN	OMP OR TOLF OR CMLB OR COA OR CRY.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12;				
RC	MEDLINE=83090452; PubMed=6294623;				
RA	Inokuchi K., Mutoh N., Matsuyama S.-I., Mizushima S.;				
RT	*Primary structure of the ompf gene that codes for a major outer membrane protein of Escherichia coli K-12.*;				
RL	Nucleic Acids Res. 10:6957-6968(1982).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12 / MG1655;				
RC	MEDLINE=97426617; PubMed=9278503;				
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;				
RT	*The complete genome sequence of Escherichia coli K-12.*;				
RL	Science 277:1453-1474(1997).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12;				
RC	MEDLINE=97061202; PubMed=8905232;				
RA	Oshima T., Alpha H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horikuchi T.;				
RT	*A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.*;				
RL	DNA Res. 3:137-155(1996).				
RN	[4]				
RP	SEQUENCE OF 1-37 FROM N.A.				
RC	MEDLINE=82139379; PubMed=7037455;				
RA	Mutoh N., Inokuchi K., Mizushima S.-I.;				
RT	*Amino acid sequence of the signal peptide of OmpF, a major outer membrane protein of Escherichia coli.*;				
RL	FEBS Lett. 137:171-174(1982).				
RN	[5]				
RP	SEQUENCE OF 23-362.				
RC	MEDLINE=82256494; PubMed=7049161;				
RA	Chen R., Kramer C., Schmidmayr W., Chen-Schmeisser U., Henning U.;				

RT "Primary structure of major outer-membrane protein I (ompF protein,
RT porin) of Escherichia coli B/r.";
RL Biochem. J. 203:33-43(1982).
RN [6]
RP SEQUENCE OF 33-63 FROM N.A.
RX MEDLINE-86033642; PubMed-2997131;
RA Nogami T., Mizuno T., Mizushima S.;
RT "Construction of a series of ompF-ompC chimeric genes by in vivo
RT homologous recombination in Escherichia coli and characterization of
RT the translational products.";
RL J. Bacteriol. 164:797-801(1985).
RN [7]
RP SEQUENCE OF 23-34 AND 39-47.
RX STRAIN-K12 / EMG2;
RC MEDLINE-97443975; PubMed-9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [8]
RP SEQUENCE OF 23-27.
RX STRAIN-K12 / W3110;
RC MEDLINE-98291876; PubMed-9629924;
RA Mollay M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,
RT Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;
RT "Extraction of membrane proteins by differential solubilization for
RT separation using two-dimensional gel electrophoresis.";
RL Electrophoresis 19:837-844(1998).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE-92375189; PubMed-1380671;
RA Cowan S.W., Schlirner T., Rummel G., Stelert M., Ghosh R.,
RT Pauplit R.A., Jansonsius J.N., Rosenbusch J.P.;
RT "Crystal structures explain functional properties of two E. coli
RT porins.";
RL Nature 358:727-733(1992).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF MUTANT ASP-141.
RX MEDLINE-95024177; PubMed-7524100;
RA Jeanteur D., Schlirner T., Fourrel D., Simonet V., Rummel G., Wildmer C.,
RT Rosenbusch J.P., Pattus F., Pages J.M.;
RT "Structural and functional alterations of a colicin-resistant mutant
RT of OmpF porin from Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:10675-10679(1994).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE-99060042; PubMed-9843370;
RA Phale P.S., Philippsen A., Kiehnhaber T., Koehnle R., Phale V.P.,
RT Schlirner T., Rosenbusch J.P.;
RT "Stability of trimeric OmpF porin: the contributions of the latching
RT loop 12.";
RL Biochemistry 37:15663-15670(1998).
RN [1-1]
RP FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW
CC SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER
CC MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2.
CC [1-1] SUBUNIT: HOMOTRIMER.
CC [1-1] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC [1-1] SIMILARITY: BELONGS TO THE OMP/POH FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; J01655; AAA2424.1; -;
CC EMBL; AE000195; AAC74015.1; -;
CC EMBL; D90730; BAA35675.1; -;
CC EMBL; D90731; BAA35681.1; -;
CC PIR; A03431; MMECF.
CC PIR; A25029; A25029.

DR PDB; 2OMF; 07-DEC-95.
DR PDB; 1OPF; 07-FEB-95.
DR PDB; 1GFV; 07-DEC-96.
DR PDB; 1GFN; 07-DEC-96.
DR PDB; 1GFO; 07-DEC-96.
DR PDB; 1GFR; 07-DEC-96.
DR PDB; 1GRO; 07-DEC-96.
DR PDB; 1MPF; 07-FEB-95.
DR PDB; 1RT9; 13-JAN-99.
DR SWISS-2DPAGE; P02931; COLI.
DR ECO2DBASE; B036.0; 6TH EDITION.
DR EcogenBase; EG10671; ompF.
DR InterPro; IPR001702; -;
DR InterPro; IPR001897; -;
DR Pfam; PF00267; Gram-ve-porins; 1.
DR PRINTS; PR00182; ECOLIIPORIN.
DR PRINTS; PR00183; ECOLIIPORIN.
DR POSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; signal;
KW 3D-structure; 1
KW 3D-structure; 1
KW CHAIN; 22
KW CONFLICT; 88
KW CONFLICT; 139
KW CONFLICT; 284
KW CONFLICT; 284
KW STRAND; 24
KW TURN; 29
KW TURN; 30
KW TURN; 31
KW TURN; 45
KW TURN; 50
KW STRAND; 53
KW STRAND; 58
KW STRAND; 62
KW STRAND; 77
KW TURN; 95
KW TURN; 102
KW TURN; 113
KW STRAND; 116
KW TURN; 126
KW TURN; 127
KW HELIX; 126
KW TURN; 132
KW TURN; 143
KW TURN; 147
KW TURN; 150
KW STRAND; 154
KW HELIX; 165
KW TURN; 171
KW STRAND; 173
KW STRAND; 183
KW TURN; 188
KW STRAND; 192
KW TURN; 195
KW TURN; 205
KW TURN; 207
KW TURN; 217
KW HELIX; 220
KW TURN; 224
KW STRAND; 227
KW STRAND; 232
KW TURN; 244
KW TURN; 245
KW STRAND; 247
KW STRAND; 257
KW TURN; 261
KW TURN; 265
KW STRAND; 268
KW TURN; 272
KW STRAND; 275
KW TURN; 288
KW STRAND; 291
KW TURN; 306
KW STRAND; 309
KW STRAND; 329
KW STRAND; 338
KW STRAND; 351
KW STRAND; 351
KW STRAND; 353
SQ SEQUENCE 362 AA; 39333 MW; 3F0974D96DB65464 CRC64;

Query Match 99.8%; Score 1805; DB 1; Length 362;
 Best Local Similarity 100.0%; Pred. No. 1.7e-121;
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEIYNKDGKNDVLYGKAVGLHFFSKNGENSGYSGNDMTYARLGFKEGTQINSDLTGXG 60
 |||||||
 DB 23 AEIYNKDGKNDVLYGKAVGLHFFSKNGENSGYSGNDMTYARLGFKEGTQINSDLTGXG 82
 |||||||
 QY 61 WEYNFQGNSEGDADQGNKTRFLAFLAKTADVGSFDYGRNRYVYDALGYTDMLEPFGG 120
 |||||||
 DB 83 WEYNFQGNSEGDADQGNKTRFLAFLAKTADVGSFDYGRNRYVYDALGYTDMLEPFGG 142
 |||||||
 QY 121 DTYASDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNKERTARNSDGGGSSY 180
 |||||||
 DB 143 DTYASDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNKERTARNSDGGGSSY 202
 |||||||
 QY 181 EYEGFIVGAVGADRTNLQEAQPLGNGKRAEQWATGLKYDANNIYLAANGETRNATPI 240
 |||||||
 DB 203 EYEGFIVGAVGADRTNLQEAQPLGNGKRAEQWATGLKYDANNIYLAANGETRNATPI 262
 |||||||
 QY 241 TNKFTNTSGRANKTQDVLVAQVQDFGLRPSIAYTKSKAKDVEGIGDVLVNFVGAT 300
 |||||||
 DB 263 TNKFTNTSGRANKTQDVLVAQVQDFGLRPSIAYTKSKAKDVEGIGDVLVNFVGAT 322
 |||||||
 QY 301 YFENKMSYVDYTIINQIDSDNKLGVGSDPTVAAGIYQF 340
 |||||||
 DB 323 YFENKMSYVDYTIINQIDSDNKLGVGSDPTVAAGIYQF 362
 |||||||

RESULT 2
 PHOE_ECOLI STANDARD; PRT: 351 AA.
 ID P02932;
 AC P02932;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
 GN PHOE OR OMP.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=83189086; Pubmed=6341601;
 RA Overbeek N., Bergmans H., van Mansfeld F., Lugtenberg B.;
 RT "Complete nucleotide sequence of phoe, the structural gene for the
 RT phosphate limitation inducible outer membrane pore protein of
 RT Escherichia coli K12.";
 RL J. Mol. Biol. 163:513-532(1983).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RA Takemoto K., Mori H., Mureyama N., Katoko K., Yano M., Itoh T.,
 RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
 RA Mizuno T., Makino K., Nakata A., Yura T., Sempel G., Mizobuchi K.;
 RT "Systematic sequencing of the Escherichia coli genome: Analysis of the
 RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=9742617; Pubmed=9278503;
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RC SEQUENCE FROM N.A.
 RP Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,

RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurli O.,
 RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RC SEQUENCE OF 1-22 FROM N.A.
 RX MEDLINE=84297233; Pubmed=6089111;
 RA Deutch A.H., Rushlow K.E., Smith C.J.;
 RT "Analysis of the Escherichia coli probA locus by DNA and protein
 RT sequencing.";
 RL Nucleic Acids Res. 12:6337-6355(1984).
 RN [6]
 RC MUTAGENESIS OF PHE-351.
 RX MEDLINE=91162638; Pubmed=1848301;
 RA Struyve M., Moons M., Tommassen J.;
 RT "Carboxy-terminal phenylalanine is essential for the correct assembly
 RT of a bacterial outer membrane protein.";
 RL J. Mol. Biol. 218:141-148(1991).
 RN [7]
 RC X-RAY CRYSTALLOGRAPHY (6.0 ANGSTROMS).
 RX MEDLINE=91172301; Pubmed=1848682;
 RA Jap B.K., Malian P.J., Gehring K.;
 RT "Structural architecture of an outer-membrane channel as determined
 RT by electron crystallography.";
 RL Nature 350:167-170(1991).
 RN [8]
 RC X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE=92375189; Pubmed=1380671;
 RA Cowan S.W., Schirmer T., Rummel G., Stelert M., Ghosh R.,
 RA Paupit R.A., Jansonsius J.N., Rosenbusch J.P.;
 RT "Crystal structures explain functional properties of two E. coli
 RT porins.";
 RL Nature 358:727-733(1992).
 RN [9]
 RC TOPOLOGY.
 RX MEDLINE=93172954; Pubmed=7679770;
 RA Struyve M., Visser J., Adriaanse H., Benz R., Tommassen J.;
 RT "Topology of Phoe porin: the 'eyelet' region.";
 RL Mol. Microbiol. 7:131-140(1993).
 CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN E. COLI CELLS
 CC ARE GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
 CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
 CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
 CC SOLUTES.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: V00316; CA23605.1; -;
 CC EMBL: D83536; BA77910.1; -;
 CC EMBL: AE000132; AAC73345.1; -;
 CC EMBL: U70214; AAB08661.1; -;
 CC EMBL: X00786; CA25362.1; -;
 CC PIR: A03432; MMECP.
 CC PDB: 1PHO; 3I-OCT-93.
 CC SWISS-2DPAGE: P02932; COLI.
 CC ECO2DBASE: B037.0; 6TH EDITION.
 CC EcoGene: EG10729; PHOE.
 CC InterPro: IPR001702; -;
 CC InterPro: IPR001897; -;
 CC Pfam: PF00267; Gram-ve.-porins; 1.
 CC PRINTS: PR00182; ECOLIPORIN.
 CC PRINTS: PR00183; ECOLIPORIN.
 CC PROSITE: PS00576; GRAM-NEG. PORIN; 1.
 KW Outer membrane; Transmembrane; Porin; Signal; 3D-structure.

```

FT SIGNAL 1 21
FT CHAIN 22 351
FT MUTAGEN 351 351
FT STRAND 25 26
FT STRAND 30 35
FT STRAND 37 45
FT STRAND 52 53
FT STRAND 56 65
FT STRAND 71 82
FT STRAND 87 88
FT STRAND 93 104
FT TURB 105 107
FT STRAND 108 116
FT TURB 119 119
FT HELIX 120 123
FT TURB 124 126
FT TURB 140 141
FT TURB 143 145
FT STRAND 147 157
FT TURB 159 162
FT STRAND 166 173
FT STRAND 176 176
FT TURB 182 183
FT STRAND 185 185
FT STRAND 188 197
FT TURB 198 201
FT STRAND 202 212
FT HELIX 215 217
FT TURB 218 219
FT STRAND 227 239
FT TURB 240 241
FT STRAND 242 252
FT STRAND 264 274
FT STRAND 280 292
FT STRAND 294 294
FT TURB 295 297
FT STRAND 298 298
FT STRAND 301 313
FT STRAND 318 327
FT TURB 334 335
FT STRAND 342 350
SQ SEQUENCE 351 AA: 38922 MW: 249E2E362C030C9A CRC64:

Query Match 61.4%; Score 1110; DB 1; Length 351;
Best Local Similarity 62.1%; Pred. No. 3.9e-72;
Matches 213; Conservative 38; Mismatches 76; Indels 16; Gaps 5;

```

```

RESULT 3
PHOE_ENTCL
ID PHOE_ENTCL STANDARD; PRT: 350 AA.
AC 047490;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
GN PHOE.
OS Enterobacter cloacae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87190417; PubMed=3032618;
RA van der Ley P., Bekkers A., van Meersbergen J., Tommassen J.;
RT "A comparative study on the phoE genes of three enterobacterial
RT species. Implications for structure-function relationships in a
RT pore-forming protein of the outer membrane.";
RL Eur. J. Biochem. 164:469-475(1987).
CC -! FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICIPATED EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -! SUBUNIT: HOMOTRIMER.
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -! SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
-----
DR EMBL; M28296; AAA24809.1; -.
DR InterPro; IPR001702; -.
DR InterPro; IPR001897; -.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLIIPORIN.
DR PRINTS; PR00183; ECOLIIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 350
FT SEQUENCE 350 AA: 38511 MW: 101655FEBC3AFAD CRC64:

Query Match 59.3%; Score 1072.5; DB 1; Length 350;
Best Local Similarity 60.3%; Pred. No. 1.8e-69;
Matches 207; Conservative 40; Mismatches 79; Indels 17; Gaps 5;

```



```

Db 254 TPI-----SGCFANKQNFVNAQYQDFGLRPSLGVQSGKQNGEIGEDLDYKIDV 307
OY 298 GATTFYFNKMSSTYYDIINQIDSNKLGVSDDTVANGIYQF 340
Db 308 GATTFYFNKMSAFVDYKINQIDIDNKLGVSSDDIYAVGMTQF 350

RESULT 4
OMPF_SERMA
ID OMPF_SERMA STANDARD; PRT; 374 AA.
AC 033980;
DC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
GN OMPF.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UOC-51;
RX MEDLINE=97419518; PubMed=9274033;
RA Hults J.A.M., Morobec E.A.;
RT "Molecular characterization of the Serratia marcescens OmpF porin,
  and analysis of S. marcescens OmpF and OmpC osmoregulation.";
RL Microbiology 143:2797-2806(1997).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPF/PHO FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U81967; AAB69103.1; ALT_INIT.
DR InterPro: IPR001897; -.
DR Pfam: PF00267; Gram-ve_porins. 1.
DR PRINTS: PR00182; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN. 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 374 OUTER MEMBRANE PROTEIN F.
SQ SEQUENCE 374 AA; 41184 MW; BF821D385E441049 CRC64;

Query Match 59.3%; Score 1072; DB 1; Length 374;
Best Local Similarity 60.1%; Pred. No. 2,1e-6;
Matches 21; Conservative 38; Mismatches 81; Indels 24; Gaps 10;
OY 1 AEIYNKDGKNDKLVGKAVGLHYFSKNGENSGYSGNDMTYARLFGKGTQINSDLTGYGQ 60
Db 23 AEIYNKDGKNDKLVGKAVGLHYFSKNGENSGYSGNDMTYARLFGKGTQINSDLTGYGQ 77

OY 61 WEYNFQGNSESGADQAGCNKRLAFAGLKYADVGSFDYGRNYGVVYDALGYTDMLPFEGG 120
Db 78 WEYNVQSNHSAQSGTE-GTKTRLGFAKLKADYGSFDYGRNYGVLYDEGWTMLPEFGG 136

OY 121 DT-AYSDFEFGRGVAVTYRNSNPEGLVNDLNFAYQLGNEBD--TARSNDGVCVGS 177
Db 137 DTYVSDPEMTGRNGVATYNNNPFGLVNDLNFALQYQGNQNDGRDVKRQNGDGMGWS 196

OY 178 ISYEY-EGFGIVAGYAGADRNLQEAQPLGNGKRAEQMATAGLKYDANNIYLAANGFERN 236
Db 197 STYIDGEGVSGAAYASSNRDQKLRSENGKADAMTVGAKKDANNVYLAAMYAEIRN 256

```

```

OY 237 ATP1-TNKTNT-----SGFANKTODVLVAQYQDFGLRPSLGYATYKSKAK--DVEGI 286
Db 257 MTFPGGGFTNTICATENCQGGFASQTONFEVTAQYQDFGLRPEVSYLQSGKLNLPGV 316

OY 287 G-DVDLVNPFVAGATYYENKMSSTYYDIINQIDSDN--RLGVGSDDTVAVGIYQF 340
Db 317 GSDQDLVAVSVGTTYENKMSSTYYDKINLDDNDFTKATGATDIDVGVGLYQF 374

RESULT 5
PHOE_CITFR
ID PHOE_CITFR STANDARD; PRT; 351 AA.
AC 001605;
DC 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
GN PHOE.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93146376; PubMed=1337052;
RA Sierings G., Ockhuijsen C., Hoisra H., Tommassen J.;
RT "Characterization of the Citrobacter freundii phoE gene and
  development of C. freundii-specific oligonucleotides.";
RL FEMS Microbiol. Lett. 78:199-204(1992).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDICED WHEN CELLS ARE
  GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
  PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
  PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
  SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPF/PHO FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X68021; CAA48162.1; -.
DR PIR: S25520; S25520.
DR HSP: P02932; IPHO.
DR InterPro: IPR001702; -.
DR InterPro: IPR001897; -.
DR Pfam: PF00267; Gram-ve_porins. 1.
DR PRINTS: PR00182; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN. 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
SQ SEQUENCE 351 AA; 38807 MW; 4E9C144070381B5A CRC64;

Query Match 59.1%; Score 1070; DB 1; Length 351;
Best Local Similarity 59.2%; Pred. No. 2,7e-6;
Matches 203; Conservative 48; Mismatches 76; Indels 16; Gaps 5;
OY 1 AEIYNKDGKNDKLVGKAVGLHYFSKNGENSGYSGNDMTYARLFGKGTQINSDLTGYGQ 60
Db 22 AEIYNKDGKNDKLVGKAVGLHYFSKNGENSGYSGNDMTYARLFGKGTQINSDLTGIGR 76

OY 61 WEYNFQGNSESGADQAGCNKRLAFAGLKYADVGSFDYGRNYGVVYDALGYTDMLPFEGG 120
Db 77 WEAFRAGKKAQSDSQ--QKTRLAFAGSKLNLSGFDYGRNLYGALYDVAETDHPFEGG 134

```

[illegible]

Best Local Similarity 58.0%, Pred. No. 1.9e-68;
Matches 199; Conservative 46; Mismatches 82; Indels 16; Gaps

```
OY      1 AEIYNKDGKKVLDYSKAVGLHFSGKGENSESYGNGDMTARLFGKEITNSDLTGQ 60  
          ||::||::|::||::   :|| :: :    :|| :: |::::|::|::|  
Db       21 ABEYNNKNKGLVDYGKVAHMHMSDYDSK-----DGDOSYRFEFGEQTINDOLTYGR 75  
  
OY      61 WEINPGGNSSGCADNOJTKPKPLAFAGLKADVGSFDVGRRGVVVVALGYDTMLPEFG 120  
          || :: |:: |:: |::|::|::|::|::|::|::|::|::|::|::|::|:  
Db       76 WEEFSNKAEDSDSO--OKTRLAFLAGKLKDLSFPDGRNLALYLVEAMTWMPFFBG 133  
  
OY     121 D-PAYSDDFEVRGGCVATYRSNFEGVLDELINFAYOLTGENHDTPARRSNDGVGSIS 179  
          |: :|::|:|:|:||||::|::|::|::|::|::|::|::|::|::|::|:  
Db     134 DSAAQIDNMTRRASLTARYRTDPFGIVDDGLTLIOQGKNEEDRDVKQONGDGFHSVS 193  
  
OY     180 YEXEG--FGIVAGAADRFTNLQEAPLGSKGAQOMATCLKTDANNITYLAANTGERNA 237  
          ::: : |:: |:: |::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     194 YDEGSDFAVSGAYTLSDRTLREONTQRGTGDKAEMAWTVGVKYDAANIYTATPSEIRMA 253  
  
OY     238 TPTTNKTSTSGANKTOOVLLVAQYORDEFGLRPDIATKSADVDIGVDLVNYEFV 297  
          ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     254 TPV-----SGGANNTTEAFEVIDYPDFGRPSGLGYLVLSKGDIECGVEDLVNIIDV 307  
  
OY     298 GATYYFNKMNSTYVDYIIQNIDSNKLGVSDDTFPAVGVIQVF 340  
Db     308 GATYYFNKMNSAFPVDKIKINGLDSDMTLGINDDLYAIGLTQVF 350  
  
RESULT 7  
OMP_C_ECOLI STANDARD: PRT; 367 AA.  
AC p06986;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE OUTER MEMBRANE PROTEIN C PRECURSOR (PORIN OMC) (OUTER MEMBRANE  
  DE PROTEIN 1B).  
CS OMC OR MECA OR PAR.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI_TaxId=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83213433; PubMed=6304064;  
RA Mizuno T., Chou M.-Y., Inouye M.;  
  RA "A comparative study on the genes for three porins of the Escherichia  
  RP colt outer membrane. DNA sequence of the osmoregulated ompC gene.";  
  RL J Biol. Chem. 258:6932-6940(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; Pubmed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.D., Rode N.T., Burland V.,  
  RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
  RA Gregor U.J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
  RA Mau B., Shao Y.;  
  RN "The complete genome sequence of Escherichia coli K-12."  
  RT science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=97251358; Pubmed=9097040;  
RA Itoh T., Alha H., Baba T., Fujita K., Hayashi K., Inada T.,  
  RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
  RA Makino K., Miki T., Mizubuchi K., Mori H., Mori T., Motomura K.,  
  RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,  
  RA Salto N., Sampei G., Seki Y., Siyasundaram S., Tagami T.,  
  RA Tanaka J., Takemoto K., Wada C., Yamamoto Y., Horiiuch T.;  
  RN "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
  RT corresponding to the 40.1-50.0 min region on the linkage map.";  
  RL DNA Res. 3:379-392(1996).  
RN [4]
```

RP SEQUENCE OF 218-367 FROM N.A.
 RC STRAIN-K12 / BHB2600.
 RA Richterich P., Lacey N., Gryan G., Jaehn L., Mintz L., Robison K.,
 RL Church G.M.;
 RN submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 (15)
 RP SEQUENCE OF 1-22 FROM N.A., AND SEQUENCE OF 22-40.
 RX MEDLINE-83132326; PubMed-6297988;
 RA Mizuno T., Chou M.-Y., Inouye M.;
 RT "DNA sequence of the promoter region of the ompC gene and the amino
 RT acid sequence of the signal peptide of pro-OmpC protein of
 RT Escherichia coli.";
 RL Peps Lett. 151:159-164(1983).
 (16)
 RP SEQUENCE OF 32-57 FROM N.A.
 RX MEDLINE-86033642; PubMed-2997131;
 RA Nogami T., Mizuno T., Mizushima S.;
 RT "Construction of a series of ompC-ompC chimeric genes by in vivo
 RT homologous recombination in Escherichia coli and characterization of
 RT the translational products.";
 RL J. Bacteriol. 164:797-801(1985).
 (17)
 RP SEQUENCE OF 22-30.
 RC STRAIN-K12 / EMG2;
 RX MEDLINE-97443975; PubMed-9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 (18)
 RP SEQUENCE OF 22-26.
 RC STRAIN-K12 / W3110;
 RX MEDLINE-98291876; PubMed-9629924;
 RA Molloy B.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,
 RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Goley A.A.;
 RT "Extraction of membrane proteins by differential solubilization for
 RT separation using two-dimensional gel electrophoresis.";
 RL Electrophoresis 19:837-844(1998).
 CC -1- FUNCTION: FORM PASSIVE DIFFUSION PORES WHICH ALLOW SMALL MOLECULAR
 CC WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER MEMBRANE.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: K00541; AAA24243.1; -
 DR EMBL: AE000310; AAC75275.1; -
 DR EMBL: D90850; BAA15998.1; -
 DR EMBL: 000008; AAA16412.1; -
 DR PIR: A20867; MMECP.
 DR PIR: B25029; B25029.
 DR PIR: A18885; A18885.
 DR HSSP: P02931; 1GFM.
 DR ECODBASE: A035.5; 6TH EDITION.
 DR Ecogene; EG10670; ompC.
 DR InterPro: IPR001702; -
 DR InterPro: IPR001897; -
 DR Pfam: PF00267; Gram-ve_porins; 1.
 DR PRINTS: PR00182; ECOLEIPORIN.
 DR PRINTS: PR00183; ECOLEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 367 OUTER MEMBRANE PROTEIN C.
 SO SEQUENCE 367 AA; 40368 MW; 6AA9370CC8A1A225 CRC64;

Query Match 58.1%; Score 1051; DB 1; Length 367;
 Best Local Similarity 58.9%; Pred. No. 6,4e-68;
 Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
 QY 1 AEIYNQGNKVDLYGKAYGKHYFSKNGENSYGNGDMTYARLGFKEGTQINSIDTGYGQ 60
 DB 22 AEIYNQGNKVDLYGKAYGKHYFSKNGENSYGNGDMTYARLGFKEGTQINSIDTGYGQ 76
 QY 61 WEYNFGNNSEGAADQGTNKLAFAGLKYADYSEDFYGRNNGYVYDALGYTDLMPFEG 120
 DB 77 WEYQIGNSAEN--ENMSVTRAFAGLKFQDYGSPFYGRNNGYVYDWTDLPEFG 133
 QY 121 DTAYSDPEFYGRYGVATYNSNPFGLVGLNFAVQYLGKNERDT-----A 166
 DB 134 DTYSQDNFMQQRNGFATYNTDFEGLVGLNFAVQYGRNNGPSCGFTSGVTNNGR 193
 QY 167 RRSNGDVGGSISYEYEGFIVGAYGAADRTNLOE-AQPGNGKKAQMATGLKYDANNI 225
 DB 194 LRQNGDVGGSITIDYEGFIVGAYGAADRTNLOE-AQPGNGKKAQMATGLKYDANNI 253
 QY 226 YLAANYGETRNATPTNKTNTSGFANKTODVLLVAQYDFGLRPSIAYTKSRANDY-E 284
 DB 254 YLAQYQTGTNATRV-----GSLGMANKAQNFQAVQYDFGLRPSIAYTKSRANDY-E 308
 QY 285 GICDVLVNTFEVGAATYENKKNSTYVDYIINOIDSNKL-----GVSDTVAVGIYQF 340
 DB 309 GYDEDELKIVDYGATYFNKNMSTYVDYKINLID-DNQETPDAGINTDVIAGLYQF 367
 RESULT 8
 ID PHOE_KLEPN STANDARD; PRT; 351 AA.
 AC P30704;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
 GN PHOE.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87190417; PubMed-3032618;
 RA van der Ley P., Bekkers A., van Meersbergen J., Tommassen J.;
 RT "A comparative study on the pho genes of three enterobacterial
 RT species. Implications for structure-function relationships in a
 RT pore-forming protein of the outer membrane.";
 RL Eur. J. Biochem. 164:469-475(1987).
 CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
 CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
 CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
 CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
 CC SOLUTES.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M28295; AAA25121.1; -
 DR HSSP: P02932; IPHO.
 DR InterPro: IPR001702; -
 DR InterPro: IPR001897; -
 DR Pfam: PF00267; Gram-ve_porins; 1.

DR PRINTS: PR00182; ECOLI:PORIN.
DR PRINTS: PR00183; ECOLI:PORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
SQ SEQUENCE 351 AA; 38894 MW; 86A5286C12502EC7 CMC64;

Query Match 58.0%; Score 1050; DB 1; Length 351;
Best Local Similarity 58.6%; Pred. No. 7.2e-68;
Matches 201; Conservative 43; Mismatches 83; Indels 16; Gaps 5;

QY 1 AEIYKDKNKVDLYGKAVGLHYFSKNGENSTYGGDMTYARLGKGETQINSIDLTGYGQ 60
DB 22 AEVYKRNKANKLDYVGKIKAMHYFSDDYDQ-----DGDQTYVAFGKGETQINEDLTGYGR 76
QY 61 WEYNQGNNSGADAOQTNKTRIAFAGLKYADVGSFVDGRNRYVYDALGYTDMLEPGG 120
DB 77 WESEFSGKRTESDSSQ--QKTRLAFAGVKLNKNGSGFDYGRNGLYDVEAWTDMPEFGG 134
QY 121 D-TAYSDFEYGRVGVATYRNSNFFGLVDGLNFAVQYLGKNERPTARRNSGDDGVGSIS 179
DB 135 DSSAQTDMFMKRAAGLATYRNTDFGLVDGLDLYOGKNGEAKKQNGDGYGTSLIS 194
QY 180 YEYEG--FGIVGAYGAADRTNLQEAQPLGNGKKAQOMATGLKYDANNIYLAANGETRNA 237
DB 195 YDFGSTDFAVSAAYTSRDTNDQNLARAQSGSKAEAMATGLKYDANNIYLAAMYSETRKM 254
QY 238 TPITNKFTNTSGFAKKTQDVLVLAQYDFGLRPSIATYKSKAKVEGIDVDLYNREY 297
DB 255 TPT-----SGGFKAQKNFEAVAYQDFGLRPSLGLVYSGKQIEGVSIEDLVNYIDV 308
QY 298 GATYFFNKMSTYVDYIINQIDSDNKLGVSDDTVAAGIVYQF 340
DB 309 GUTYFFNKNMNAFVDYKINQKSDMKLGINDDIYALGMYQF 351

RESULT 9
PORI_BPPA2 STANDARD; PRT; 365 AA.
AC 007238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORIN PROTEIN LC PRECURSOR.
GN LC.
OS Bacteriophage PA-2.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxID=10738;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304457; PubMed=3017988;
RA Blassband A.J., Marcotte W.R. Jr., Schnaltman C.A.;
RT "Structure of the lc and mmpc outer membrane porin protein genes of
RT lambdaoid bacteriophage."
RL J. Biol. Chem. 261:12723-12732(1986).
CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER
CC THAN 4000-6000 DALTONS).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
CC MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
CC THE LYSOGENIC STATE. IN E. COLI THE EXPRESSION OF OMPC AND OMPF
CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHO FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J02580; AAA32301.1; -.
DR PIR: D25647; MMAPR2.
DR HSSP: P02931; IGFN.
DR InterPro: IPR001702; -.
DR InterPro: IPR001897; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLI:PORIN.
DR PRINTS: PR00183; ECOLI:PORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23
FT CHAIN 24 365 OUTER MEMBRANE PORE PROTEIN LC.
FT CONFLICT 99 99 H -> R (IN AA SEQUENCE).
SQ SEQUENCE 365 AA; 40290 MW; 0FBC0531FB9C0205 CMC64;

Query Match 57.8%; Score 1046; DB 1; Length 365;
Best Local Similarity 59.5%; Pred. No. 1.4e-67;
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;

QY 1 AEIYKDKNKVDLYGKAVGLHYFSKNGENSTYGGDMTYARLGKGETQINSIDLTGYGQ 60
DB 24 AEIYKDKNKLDLYGKAVNAKHYFSSNDAD-----DGDYTYARLGKGETQINDOULTGYGQ 78
QY 61 WEYNQGNNSGADAOQTNKTRIAFAGLKYADVGSFVDGRNRYVYDALGYTDMLEPGG 120
DB 79 WEYFKGNRAE--SQGSSKDKTHLAFAGLKFQDYGSDIDGRNRYVYDALGYTDMLEPGG 137
QY 121 DT-AYSDFEYGRVGVATYRNSNFFGLVDGLNFAVQYLGKNER--PTARRNSGDDGVG 176
DB 138 DTWQTQDVFMGRTGTFADRYRNNDPFGLVLDGLNFAAQYQGNKDRSDFNYTEGNDGGEF 197
QY 177 STSYREGGIVGAYGAADRTNLQE-----AQPLGNGKKAQOMATGLKYDANNIYLAAN 230
DB 198 SATYEEGFGIGATYAKDSDRTQVNAKGVLPYEVASGKNMEVMAAGLYDANNIYLYTT 257
QY 231 YGETRNATPTNKFTNTSGFANKTQDVLVLAQYDFGLRPSIATYKSKAKVEGIDVD 290
DB 258 YSEIQNMVTFADHY-----ANKAQNFVAAYQDFGLRPSVATLQSKGDLGWGDD 312
QY 291 LVNFEVATYFFNKMSTYVDYIINQIDSDN--KLGVSDDTVAAGIVYQF 340
DB 313 LVKYVDGATYFFNKMSTYVDYKINLKDNDFTKALGVSTDIYAVAGIVYQF 365

RESULT 10
PHOE_SALTI STANDARD; PRT; 350 AA.
AC 056119;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
GN PHOE.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IMS-1;
RA Torres A., Puente J.L., Calva E.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE.
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.

[3]
 RN SEQUENCE FROM N.A., SEQUENCE OF 22-33, AND CHARACTERIZATION.
 RP MEDLINE-98317278; PubMed-9642192;
 RA Prilipov A., Pale P.S., Koebnik R., Widmer C., Rosenbusch J.P.;
 RT Identification and characterization of two quiescent porin genes,
 RL ompc and ompn, in Escherichia coli BE.";
 RL J. Bacteriol. 180:3388-3392(1998).
 CC -1- FUNCTION: NON-SPECIFIC PORIN.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPc/PHOE FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE000234; AAC74459.1; -
 CC EMBL: D90775; BAA14981.1; -
 CC EMBL: D90776; BAA14986.1; -
 CC EcoGene: EG13375; ompn.
 CC InterPro: IPR001897; -
 CC InterPro: IPR001897; -
 CC Pfam: PF00267; Gram-ve_porins; 1.
 CC PRINTS: PR00182; ECOLIPIORIN.
 CC PRINTS: PR00183; ECOLIPIORIN.
 CC Outer membrane; Transmembrane; Porin; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 377
 FT SEQUENCE 377 AA; 41220 MW; 4420DZCB9BC10F95 CRC64;
 Query Match 57.2%; Score 1035; DB 1; Length 377;
 Best Local Similarity 56.9%; Pred. No. 9e-67; Indels 38; Gaps 9;
 Matches 209; Conservative 42; Mismatches 78;
 QY 1 AEYNNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGEGQ 60
 DB AEYNNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGEGQ 76
 QY 61 WEYNNFQGNSEGAADQGNKTRLAFLAGLKADYVSGFDYGNNGYVVDALGYTMLPEFGG 120
 DB WEYNNFQGNSEGAADQGNKTRLAFLAGLKADYVSGFDYGNNGYVVDALGYTMLPEFGG 134
 QY 121 DT--AYSDPEFVGRGVATYRNSNFFGLVLDGLNFAVOYLGNKNE-----PRTAR 167
 DB DT--AYSDPEFVGRGVATYRNSNFFGLVLDGLNFAVOYLGNKNE-----PRTAR 193
 QY 135 DSYTNADNENFTGKANGVATYRNTDFGLVGLNFAVOYOGNNEGASNGOGCTNGNRD-VR 193
 DB DSYTNADNENFTGKANGVATYRNTDFGLVGLNFAVOYOGNNEGASNGOGCTNGNRD-VR 193
 QY 168 RNSGDSVGGISLEY-EGFSGIVAGYGAADRTNLQEAOPLGNGKKAQEMATGLKYDANNIT 226
 DB RNSGDSVGGISLEY-EGFSGIVAGYGAADRTNLQEAOPLGNGKKAQEMATGLKYDANNIT 253
 QY 194 HENGDMGGLSTYDLDGSGFAGAAVSSORTNQVNHNTAAGCKADKADMTGLKYDANNIT 253
 DB HENGDMGGLSTYDLDGSGFAGAAVSSORTNQVNHNTAAGCKADKADMTGLKYDANNIT 253
 QY 227 LAANGETRATATITNKFTNTSGFANKTDVLLVAOYQDFGLRPSIATYKSKAKDVE-- 284
 DB LAANGETRATATITNKFTNTSGFANKTDVLLVAOYQDFGLRPSIATYKSKAKDVE-- 310
 QY 254 LATMYSETRMNTPFGD---SDYAVANKTONFEVTAOYQDFGLRPAVSFLMSKGRDLHAA 310
 DB LATMYSETRMNTPFGD---SDYAVANKTONFEVTAOYQDFGLRPAVSFLMSKGRDLHAA 310
 QY 285 -----GIGDVLVNFVFGVATYRNSNFFGLVLDGLNFAVOYLGNKNE-----GVGSDDTVA 333
 DB -----GIGDVLVNFVFGVATYRNSNFFGLVLDGLNFAVOYLGNKNE-----GVGSDDTVA 370
 QY 311 GGDNDPAGVDDKDLVRYKADIGATYFRNKNMSTVYDKINLLDEDSFYAANGISTDDIVA 370
 DB GGDNDPAGVDDKDLVRYKADIGATYFRNKNMSTVYDKINLLDEDSFYAANGISTDDIVA 370
 QY 334 VGIYQOF 340
 DB VGIYQOF 377
 RESULT 14
 OMPF_SALTI STANDARD; PRT; 363 AA.
 ID OMPF_SALTI
 AC 056113;
 DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE OUTER MEMBRANE PROTEIN F PRECURSOR (PORIN OMPF) (OUTER MEMBRANE
 DE PROTEIN S3).
 GN OMPF OR OMP53.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_Taxid-601;
 RN SEQUENCE FROM N.A.
 RC SPRAIN-IMSS-1;
 RA Fernandez-Mora M., Calva E.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW
 CC SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER
 CC MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2 (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPc/PHOE FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X8957; CAA61905.1; -
 CC InterPro: IPR001702; -
 CC InterPro: IPR001897; -
 CC Pfam: PF00267; Gram-ve_porins; 1.
 CC PRINTS: PR00182; ECOLIPIORIN.
 CC PRINTS: PR00183; ECOLIPIORIN.
 CC PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 CC Outer membrane; Transmembrane; Porin; Phage recognition; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 363
 FT SEQUENCE 363 AA; 40048 MW; F6069B34E9516859 CRC64;
 Query Match 57.0%; Score 1031.5; DB 1; Length 363;
 Best Local Similarity 57.6%; Pred. No. 1.5e-66;
 Matches 200; Conservative 54; Mismatches 80; Indels 13; Gaps 8;
 QY 1 AEYNNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGEGQ 60
 DB AEYNNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGEGQ 78
 QY 23 AEYNNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGEGQ 78
 DB AEYNNKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGTQINSDLTGEGQ 78
 QY 61 WEYNNFQGNSEGAADQGNKTRLAFLAGLKADYVSGFDYGNNGYVVDALGYTMLPEFGG 120
 DB WEYNNFQGNSEGAADQGNKTRLAFLAGLKADYVSGFDYGNNGYVVDALGYTMLPEFGG 137
 QY 79 WEYTRKADRADG-EGQNSNLVRLAFLAGLKADYVSGFDYGNNGYVVDALGYTMLPEFGG 137
 DB WEYTRKADRADG-EGQNSNLVRLAFLAGLKADYVSGFDYGNNGYVVDALGYTMLPEFGG 137
 QY 121 DT--AYSDEPFGVGRGVATYRNSNFFGLVLDGLNFAVOYLGNKNEPRTARRS-NGDGVGG 176
 DB DT--AYSDEPFGVGRGVATYRNSNFFGLVLDGLNFAVOYLGNKNEPRTARRS-NGDGVGG 197
 QY 138 ETMGAGATDYMYTSRAAGLTLTYRNSDFFGLVLDGLSFGIYQGNKQDNHNSINSNGDGVGY 197
 DB ETMGAGATDYMYTSRAAGLTLTYRNSDFFGLVLDGLSFGIYQGNKQDNHNSINSNGDGVGY 197
 QY 177 SISIEYFGFIVAGYGAADRTNLQEAOPLGNGKKAQEMATGLKYDANNITLANNYGETRN 236
 DB SISIEYFGFIVAGYGAADRTNLQEAOPLGNGKKAQEMATGLKYDANNITLANNYGETRN 256
 QY 198 TMAIEPFGFVYTAAYSKRTNDQDND-NGGDRAESMAVAGAKDANNVLLAAYATETRN 256
 DB TMAIEPFGFVYTAAYSKRTNDQDND-NGGDRAESMAVAGAKDANNVLLAAYATETRN 256
 QY 237 ATPITNKFTNTSGFANKTDVLLVAOYQDFGLRPSIATYKSKAKDVEGI-GVDLVNPF 295
 DB ATPITNKFTNTSGFANKTDVLLVAOYQDFGLRPSIATYKSKAKDVEGI-GVDLVNPF 316
 QY 257 MSIVENTVDTVEANKTKQULEVVAOYQDFGLRPSIATYKSKAKDVEGI-GVDLVNPF 316
 DB MSIVENTVDTVEANKTKQULEVVAOYQDFGLRPSIATYKSKAKDVEGI-GVDLVNPF 316
 QY 296 EVGATYFRNKNMSTVYDIINQDSDNKLK--VGSDDTVAVGIYQOF 340
 DB EVGATYFRNKNMSTVYDIINQDSDNKLK--VGSDDTVAVGIYQOF 363
 QY 317 QAGATYFRNKNMSTVYDIINQDSDNKLK--VGSDDTVAVGIYQOF 363
 DB QAGATYFRNKNMSTVYDIINQDSDNKLK--VGSDDTVAVGIYQOF 363
 RESULT 15

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:05:41 ; Search time 95.41 Seconds

(without alignments)
472.865 Million cell updates/sec

Title: US-09-490-291-9
Perfect score: 1809
Sequence: 1 AEIYNKDGKNDLVYKRAVGL.....NKLGVSGDDPVAVGVYQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 133305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : SPTREMBL.16.*
1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_unclassified:*
13: SP_vertebrate:*
14: SP_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1044.5	57.7	374	2	087754
2	1036	57.3	377	2	085030
3	1029.5	56.9	364	2	09K597
4	1029.5	56.9	378	2	09K3E6
5	1028.5	56.9	366	2	09RH85
6	1022.5	56.5	333	2	09RAW3
7	1022.5	56.5	333	2	09RAW2
8	1022.5	56.5	363	2	09K3E7
9	1021.5	56.4	333	2	09RR59
10	1019.5	56.4	333	2	09S613
11	1017.5	56.2	360	2	09FR89
12	934.5	51.7	359	2	087753
13	845	45.1	315	2	09EXH8
14	770	42.6	315	2	09EC71
15	428.5	23.7	130	2	052641
16	310	17.1	111	2	P94857
17	310	17.1	111	2	P94858
18	310	17.1	111	2	P94860
19	309	17.1	111	2	P94635

20	260.5	14.4	351	2	09RNA1	09nal aeromonas h
21	233.5	12.9	96	2	P77519	P77519 escherichia
22	224	12.4	331	2	09KU02	09KU02 vibrio chol
23	197.5	10.9	330	2	P94743	P94743 elkenella c
24	170.5	9.4	69	2	P76871	P76871 escherichia
25	167.5	9.3	329	2	054350	054350 pasteurella
26	162.5	9.0	333	2	09R2D4	09R2D4 pasteurella
27	162.5	9.0	333	2	054339	054339 pasteurella
28	160.5	8.9	329	2	054345	054345 pasteurella
29	159	8.8	323	2	09X586	09X586 neisseria m
30	158	8.7	323	2	066067	066067 neisseria m
31	157	8.7	361	2	045106	045106 burkholderi
32	155.5	8.6	50	2	09EC72	09EC72 yersinia pe
33	155.5	8.6	322	2	051273	051273 neisseria m
34	155.5	8.6	323	2	034343	054343 pasteurella
35	154	8.5	332	2	P72072	P72072 neisseria f
36	153.5	8.5	333	2	054344	054344 pasteurella
37	153	8.5	339	2	051947	051947 neisseria g
38	152.5	8.4	332	2	P72113	P72113 neisseria s
39	152.5	8.4	343	2	054340	054340 pasteurella
40	152	8.4	323	2	070078	070078 neisseria m
41	152	8.4	328	2	054347	054347 pasteurella
42	152	8.4	328	2	054349	054349 pasteurella
43	151.5	8.4	313	2	051266	051266 neisseria m
44	151.5	8.4	322	2	066062	066062 neisseria m
45	151	8.3	327	2	09X587	09X587 neisseria m

ALIGNMENTS

```

RESULT 1
ID 087754 PRELIMINARY; PRT; 374 AA.
AC 087754;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OMPK37 PORIN PRECURSOR.
GN OMPK37.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxId=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SD8;
RA Domenech-Sanchez A., Hernandez-Alles S., Martinez-Martinez L.,
RA Benedi V.J., Alberti S.,
RT Identification and characterization of a novel porin of Klebsiella
RT pneumoniae: its role in beta-lactam antibiotics resistance.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF011502; CA09666.1; -
DR HSSP: Q48473; 10SM.
DR InterPro: IPR000408; -
DR InterPro: IPR001702; -
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLEIPORIN.
DR PROSITE: PS00626; RCOL_2; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL. 1 21
FT CHAIN 22 374 OMPK37 PORIN.
FT SEQUENCE 374 AA; 41612 MW; B66FDD74A1A169B CRC64;

```

Query Match 57.7%; Score 1044.5; DB 2; Length 374;
Best Local Similarity 58.5%; Pred. No. 9.3e-69;
Matches 214; Conservative 37; Mismatches 76; Indels 39; Gaps 9;

OY 1 AEIYNKDGKNDLVYKRAVGLHFSKNGENSYGNGDMFYARLGFKGTGINSDLTGYYG 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 22 AEIYNKDGKNDLVYKRAVGLHFSKNGENSYGNGDMFYARLGFKGTGINSDLTGYYG 76

```

OY 61 WEYNFGNNSGADAOQTGNKTRLAFLAKYADVGSFDYGRNNGVYDALGYTDMLEPFGG 120
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
DB 77 WEYNVQANNETSSDQAM--TRLAFAIGIKYDVGSPDYGRNGVLYLVEGWTDLPEFGG 134
OY 121 DT-AVSDDFVGRGVGATYRNSNFEGLVDGLNFAVOYLKGNERTA-----166
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 135 DSTYADNFMAGRANGVATYRNSDFGLVEGLNFALOYOGKNEGQANODINVCNNSSD 194
OY 167 ---RRSNGDVGGSISYEYEGFGI--VGAYGAADRTNLQEAQPLGNGKKAQWATYGLKYD 221
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
DB 195 SDVAFDNGDGFGLSTYDF--CMGISAAAYTSSDRINDQMTNARBDKAEAWTAGLKYD 253
OY 222 ANNTYLAANGETRNAPITKFTNTSGFANKQDVLLVAQYDFGLRPSIATYKSKAK 281
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 254 ANDTYLATMYSERNNMTPYGN----DGVANKTQNFVYLAQYDFGLRPAISLYQSKGK 308
OY 282 DVEGIG---DVDLVNFEVGATYFENKMNSTYVDYIINQIDSNKL----GVGSDDTYAV 334
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 309 DLYNNGKRAKADLVKAYDVGATYFENRMSTYVDYKINLSDGNDKFYEDNGISTDNLYAL 368
OY 335 GIVYQF 340
    || : : : :
DB 369 GLVYQF 374

RESULT 2
085030 PRELIMINARY; PRT; 377 AA.
AC 085030;
DB 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PORIN OMPN.
GN OMPN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BE (BL21);
RX MEDLINE=98317276; PubMed=9642192;
RA PiliDPOV A., Phale P.S., Koebnik R., Widmer C., Rosenbusch J.P.;
RT "Identification and characterization of two quiescent porin genes,
RL nmpC and ompN, in Escherichia coli BE.";
DR EMBL: AF035618; AAC38644.1; -.
HSSP: P02931; IGPN.
DR InterPro: IPR000408; -.
DR InterPro: IPR001702; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 377 AA; 41236 MW; 7F4D681A2BC10F8C CRC64;

```

```

DB 194 HENDGWSLSTTYDGLGFSAGAAVYTSDDRTNDQVNHNTAAGGKADAWTAGLKVDANNIY 253
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
OY 227 LAANYGETRNATPTTKFTNTSGFANKQDVLLVAQYDFGLRPSIATYKSKAKVE-- 284
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 254 LATMYSETRNMTPFGD---SDYAVANKTONFEVYLAQYDFGLRPAVSFLMSKGRDLHAA 310
OY 285 -----GIDVDLVNFEVGATYFENKMNSTYVDYIINQIDSNKL----GVGSDDTVA 333
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 311 GGADNPAGVDKDLVAYDVGATYFENKMNSTYVDYKINLDEDDSYTANGISTDDIYA 370
OY 334 VGIVYQF 340
    || : : : :
DB 371 LGIVYQF 377

RESULT 3
09K597 PRELIMINARY; PRT; 364 AA.
AC 09K597;
DB 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OUTER MEMBRANE PORIN C PRECURSOR.
GN OMPC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC low A.S.; Mackenzie F.M., Gould I.M., Booth I.R.;
RA "Parallel evolution of multi-resistant bacteria in a patient with
RT recurrent septicemia: unique data that support the presence of
RT separate protected environments.";
RL submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
DR EMBL: AJ295721; CAC01403.1; -.
DR InterPro: IPR000504; -.
DR InterPro: IPR001702; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
DR Outer membrane; porin; Signal; Transmembrane.
FT SIGNAL 1
FT POTENTIAL 21
SQ SEQUENCE 364 AA; 40312 MW; 9B583F2C11344E31 CRC64;

```

```

Query Match 57.3%, Score 1036; DB 2; Length 377;
Best Local Similarity 57.2%, Pred. No. 3.9e-68;
Matches 210; Conservative 41; Mismatches 78; Indels 38; Gaps 9;
OY 1 AEIYKDKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGFTQINSDLTGYGQ 60
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
DB 22 AEIYKDKGNKVDLYGKAVGLHYFSNDSAK-----DGDOSTALGLRGKETQINDQLTGYGQ 76
OY 61 WEYNFGNNSGADAOQTGNKTRLAFLAKYADVGSFDYGRNNGVYDALGYTDMLEPFGG 120
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
DB 77 WEYNQANNETSSDQAM--TRLAFAIGIKYDVGSPDYGRNGVLYLVEGWTDLPEFGG 134
OY 121 DT-AVSDDFVGRGVGATYRNSNFEGLVDGLNFAVOYLKGN-----RDYAR 167
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 135 DSTYADNFMAGRANGVATYRNSDFGLVEGLNFAVOYOGKNEGASNGOEGTNGRD--VR 193
OY 168 RNSGDSGVSISYER--EGRGIVGAYGAADRTNLQEAQPLGNGKKAQWATYGLKYDANNIY 226

```

```

Query Match 56.9%, Score 1029.5; DB 2; Length 364;
Best Local Similarity 58.5%, Pred. No. 1.1e-67;
Matches 209; Conservative 40; Mismatches 77; Indels 31; Gaps 8;
OY 1 AEIYKDKGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFKGFTQINSDLTGYGQ 60
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
DB 22 AEIYKDKGNKVDLYGKAVGLHYFSNDSDK-----DGDYTNHRLGFKGFTQYDQLTGYGQ 76
OY 61 WEYNFGNNSGADAOQTGNKTRLAFLAKYADVGSFDYGRNNGVYDALGYTDMLEPFGG 120
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
DB 77 WEYQYQNEPE---SDNSWTFVAFAGLKFQDVGSFDYGRNNGVYDVTSWTDYLPFGG 133
OY 121 DTAYSDDFVGRGVGATYRNSNFEGLVDGLNFAVOYLK-----NERTARRS 169
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
DB 134 DTYSDNDNMQDNGNFAVYRNTDFGLVDGLFAVOYOGKNGSAHGEQNTNGRDVEEQ 193
OY 170 NGDVGGSISYEYEGFGIYAGYGAADR--NLQEAQPLGNGKKAQWATYGLKYDANNIY 228
    ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : : ||||| : : : :
DB 194 NGDVGGSITTYRBEFGICAANVSSKRTWDDQNTTGLIGDGAELTYTGLKDYDANNIY 253
OY 229 ANYGETRNATPTTKFTNTSGFANKQDVLLVAQYDFGLRPSIATYKSKAKVD--EGIG 287
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :

```

Db 254 AAYTQYTNATRV-----GSLGWANKAQNFEAVAYQYDFGLRPSLAVLQSKGNKLGRCYD 308

OY 288 DVDLVNFEVGGATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

Db 309 DEBLKAVDVGATYFNKNMSTYVDYKINLDD-DNRFTRDAGINTDIDVALGLVYQF 364

RESULT 4

OY 09K3E6 PRELIMINARY; PRT; 378 AA.

AC 09K3E6;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE OUTER MEMBRANE PROTEIN C PRECURSOR.

GN OMP.

OS Salmonella enterica subsp. enterica serovar Minnesota.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

NCBI_Taxid=70803;

OX NCBI_Taxid=70803;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-SF111;

RA Zimmermann H., Maassenaar T.M., Laubenthaler-Preusse H., Petry F.,

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE

CC (BY SIMILARITY).

DR EMBL: Y15844; CAB96613.1; -

DR InterPro: IPR000504; -

DR InterPro: IPR001702; -

DR Pfam: PF00267; Gram-ve_porins; 1.

DR PRINTS: PR00182; ECOLNEIPORIN.

DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.

DR PROSITE: PS00030; RNP_1; UNKNOWN.1.

KW Outer membrane; Porin; Signal; Transmembrane.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 378 OUTER MEMBRANE PROTEIN C.

FT SEQUENCE 378 AA; 41269 MW; 254524EB9EC849C CRC64;

Query Match 56.9%; Score 1029.5; DB 2: Length 378;

Best Local Similarity 57.0%; Pred. No. 1.2e-67; Mismatches 78; Indels 33; Gaps 9;

Matches 208; Conservative 46; Mismatches 78; Indels 33; Gaps 9;

OY 1 AEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60

Db 22 AEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60

OY 61 WEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 120

Db 77 WEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 133

OY 121 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

Db 134 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

OY 173 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

Db 194 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

OY 228 AAYTQYTNATRV-----GSLGWANKAQNFEAVAYQYDFGLRPSLAVLQSKGNKLGRCYD 308

Db 254 AAYTQYTNATRV-----GSLGWANKAQNFEAVAYQYDFGLRPSLAVLQSKGNKLGRCYD 308

OY 284 E-----GIGVDLVNFEVGGATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

Db 314 SNGYASGYGQDIDVYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60

OY 336 IVYQF 340

Db 374 LVYQF 378

RESULT 5

OY 09R85 PRELIMINARY; PRT; 366 AA.

AC 09R85;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE OUTER MEMBRANE PROTEIN OMP.

GN OMP.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

NCBI_Taxid=562;

OX NCBI_Taxid=562;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-O157:H7;

RA Yu S.L., Syu W.J.;

RT "Altered outer membrane protein OmpC in hemorrhagic Escherichia coli

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE

CC (BY SIMILARITY).

DR EMBL: AF057355; AF21761.1; -

DR HSSP: Q48473; IOSM.

DR InterPro: IPR001702; -

DR Pfam: PF00267; Gram-ve_porins; 1.

DR PRINTS: PR00182; ECOLNEIPORIN.

DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.

KW Outer membrane; Porin; Signal; Transmembrane.

FT SEQUENCE 366 AA; 40499 MW; 6A4EAD1652565C00 CRC64;

Query Match 56.9%; Score 1028.5; DB 2: Length 366;

Best Local Similarity 58.6%; Pred. No. 1.3e-67; Mismatches 71; Indels 35; Gaps 9;

Matches 211; Conservative 43; Mismatches 71; Indels 35; Gaps 9;

OY 1 AEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60

Db 22 AEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60

OY 61 WEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 120

Db 77 WEYNNKDNVLDYKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 133

OY 121 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

Db 134 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

OY 172 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

Db 194 DTVASDPEFGVGRGVAATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

OY 230 NYGETRNATPTKFTNTSGFANKTQDVLVAOYQDFGLRPSLAVLQSKGNKLGRCYD 308

Db 253 NYGETRNATPTKFTNTSGFANKTQDVLVAOYQDFGLRPSLAVLQSKGNKLGRCYD 308

OY 288 ---DVDLVNFEVGGATYFNKNMSTYVDYIINOISDNKL----GVGSDPTVAVGIVYQF 340

Db 308 NYDEDLKAVDVGATYFNKNMSTYVDYKINLDD-DNRFTRDAGINTDIDVALGLVYQF 364

RESULT 6

OY 09RAM3 PRELIMINARY; PRT; 333 AA.

AC 09RAM3;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE PHOSPHATE PORIN (FRAGMENT).

GN PHOE.

OS Klebsiella pneumoniae (subsp. ozaenae).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=574;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1436;
 RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,
 RT "Phylogenetic analysis of Calymmatobacterium granulomatis.";
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: HOMOTIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
 (BY SIMILARITY).
 CC EMBL: AF009229; AAD21517.1; -.
 DR HSSP: P02932; 1PHO.
 DR InterPro: IPR001702; -.
 DR Pfam: PF00267; Gram-ve.porins; 1.
 DR PRINTS: PR00182; ECOLNEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Porin; Signal; Transmembrane.
 FT NON_TER 1 1
 FT SEQUENCE 333 AA; 36839 MW; 68AF347207770ABA CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 333;
 Best Local Similarity 58.9%; Pred. No. 3,2e-67;
 Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

QY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSYGGNDMTYARLGKFGTQINSDLTGYGO 60
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 15 AEYNNKANKKLDYVGKIKAMHYFSDYDSK----DGDQTVYRFGIKGETQINDLTGIGR 69
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 61 WEYNQGNNSRGADQAGTGNKTRLAFAGLKADVGSFDRGNVGVYDALGYTDLMPERGG 120
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 70 WESESGKTKESDSSQ---KTRLAFAGYKLNKYSFDRGNLGAIDYEAWTDMPERGG 126
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 121 D-TAVSDPEFVGVRGVATYRNSNFFGLVDGINFAVOYLKNERDTARRSGDVGSGIS 179
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 127 DSAQTDFNMFRRASGLATYRNTDFGLVDGLTLQYQKNEGREAKQKNDGVGTSLIS 186
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 180 YEYEG--FGIYGAGADRTNLQEAQPLNGCKKAEQWATGKLYDANNITYLANGETRNA 237
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 187 YDFGSDFAVSAAYTSDBRTDNLARGQSKAEAWATGKLYDANNITYLANGETRNA 246
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 238 TPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKVEGIGDVLVNYFEV 297
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 247 TPI-----SGGFANKAQNFEAVAOYQDFGLRPSIGYLVSKGKIDEGVSGEDLVNYIDV 300
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 298 GATYYFNKNMSTYVDYIINQIDSDNKLGVGSD 330
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 301 GLTYFFNKNMNAFVDYKINQKSDNKLGINDD 333
 ||||| :||:|||| :||| : :||| | ||||| |||||

RESULT 7
 Q9RAM2 PRELIMINARY; PRT; 333 AA.
 AC Q9RAM2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PHOSPHATE PORIN (FRAGMENT).
 GN PHE.
 OS Klebsiella pneumoniae subsp. rhinoscleromatis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=39831;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NC05046;
 MEDLINE-20023050; PubMed-10555350;
 RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,

RA Kemp D.J.;
 RT "Phylogenetic evidence for reclassification of Calymmatobacterium
 RT granulomatis as Klebsiella granulomatis comb. nov.";
 RT Int. J. Syst. Bacteriol. 49:1695-1700(1999).
 CC -1- SUBUNIT: HOMOTIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
 (BY SIMILARITY).
 CC EMBL: AF009230; AAD21518.1; -.
 DR InterPro: IPR001702; -.
 DR Pfam: PF00267; Gram-ve.porins; 1.
 DR PRINTS: PR00182; ECOLNEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Porin; Signal; Transmembrane.
 FT NON_TER 1 1
 FT SEQUENCE 333 AA; 36839 MW; 68AF347207770ABA CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 333;
 Best Local Similarity 58.9%; Pred. No. 3,2e-67;
 Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

QY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSYGGNDMTYARLGKFGTQINSDLTGYGO 60
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 15 AEYNNKANKKLDYVGKIKAMHYFSDYDSK----DGDQTVYRFGIKGETQINDLTGIGR 69
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 61 WEYNQGNNSRGADQAGTGNKTRLAFAGLKADVGSFDRGNVGVYDALGYTDLMPERGG 120
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 70 WESESGKTKESDSSQ---KTRLAFAGYKLNKYSFDRGNLGAIDYEAWTDMPERGG 126
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 121 D-TAVSDPEFVGVRGVATYRNSNFFGLVDGINFAVOYLKNERDTARRSGDVGSGIS 179
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 127 DSAQTDFNMFRRASGLATYRNTDFGLVDGLTLQYQKNEGREAKQKNDGVGTSLIS 186
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 180 YEYEG--FGIYGAGADRTNLQEAQPLNGCKKAEQWATGKLYDANNITYLANGETRNA 237
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 187 YDFGSDFAVSAAYTSDBRTDNLARGQSKAEAWATGKLYDANNITYLANGETRNA 246
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 238 TPITNKFTNTSGFANKTQDVLVLAQYQDFGLRPSIATYKSKAKVEGIGDVLVNYFEV 297
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 247 TPI-----SGGFANKAQNFEAVAOYQDFGLRPSIGYLVSKGKIDEGVSGEDLVNYIDV 300
 ||||| :||:|||| :||| : :||| | ||||| |||||
 QY 298 GATYYFNKNMSTYVDYIINQIDSDNKLGVGSD 330
 ||||| :||:|||| :||| : :||| | ||||| |||||
 DB 301 GLTYFFNKNMNAFVDYKINQKSDNKLGINDD 333
 ||||| :||:|||| :||| : :||| | ||||| |||||

RESULT 8
 Q9K3E7 PRELIMINARY; PRT; 363 AA.
 AC Q9K3E7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE OUTER MEMBRANE PROTEIN FX PRECURSOR.
 GN OMPX.
 OS Salmonella enterica subsp. enterica serovar Minnesota.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=70803;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF1111;
 RA Zimmermann H., Wassenaar T.M., Laubenheimer-Preuss H., Petry F.,
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y15843; CAB96614.1; -.
 DR InterPro: IPR001702; -.
 DR Pfam: PF00267; Gram-ve.porins; 1.
 DR PRINTS: PR00182; ECOLNEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; UNKNOWN_1.
 KM Signal.
 FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 363 OUTER MEMBRANE PROTEIN FX.
SQ SEQUENCE 363 AA: 40090 MW: F5058DEDEA516859 CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 363;
Best Local Similarity 57.3%; Pred. No. 3.6e-67;
Matches 199; Conservative 54; Mismatches 81; Indels 13; Gaps 8;

OY 1 AEIYKDKGKVDLYGKAVGLHFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 60
DB 23 AEIYKDKGKVDLYGKAVGLHFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 78
OY 61 WEYNFOGNSSEGADQGTGKTRLAFLAGLYADVSFDRGNGVVDALGYTDMLEPFG 120
DB 79 WEYRKADAEQ-EGQNSLVRLAFAGLYAEVSGIDYGRNGVVDALGYTDMLEPFG 137
OY 121 DT---AYSDFEVRGVATYRNSNFEGLVDGLNFAVOYLGNEDTRARRS-NGDVG 176
DB 138 ETWGAAYTDNMTSRAGGLTYRNSDFGLVDGLSFGLYOGNGDHSINSNGDVG 197
OY 177 SISEYEGFGIYAGVADRTNLOEAQPLGNGKAEQMATGLKYDANNITYLAANGETR 236
DB 198 TWAYEPDGFVTAAYNSKRTNDQDRD-CNGDRAESMAVGAKYDANNITYLAAYETRN 256
OY 237 ATPINKFTNTSGFANKQDVLVAOYQDFGLRPSIATYKSKAKVEGI-GEVDLVN 295
DB 257 MSIVENTYVDYEMANKTQNLLEVVAOYQDFGLRPSIATYKSKAKVEGI-GEVDLVN 316
OY 296 EVGATYFNKMNSTVYDIINOIDSQKLG--VGSDDYAVAGVYQ 340
DB 317 QGATYFNKMNSTVYDIINOIDSQKLG--VGSDDYAVAGVYQ 363
RESULT 9
OYR59 PRELIMINARY: PRT: 333 AA.
AC OYR59: 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAY-2000 (TREMREL. 16, Last annotation update)
DE PHOSPHATE PORIN (FRAGMENT).
GN PHOE.
OS Calymmatobacterium granulomatis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Calymmatobacterium.
OX NCBI_TaxID=39824;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20023050; PubMed=10555350;
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,
Kemp D.J.;
RT "Phylogenetic evidence for reclassification of Calymmatobacterium
granulomatis as Klebsiella granulomatis comb. nov.";
RL Int. J. Syst. Bacteriol. 49:1695-1700(1999).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
(BY SIMILARITY).
CC EMBL: AF009231; AAD21519.1; -.
DR HSSP: P02932; 1PHO.
DR InterPro: IPR001702; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLIPEPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT NON_TER 1
FT NON_TER 333
SQ SEQUENCE 333 AA: 36786 MW: 7CEA2F68B998B0A CRC64;

Query Match 56.5%; Score 1021.5; DB 2; Length 333;
Best Local Similarity 58.9%; Pred. No. 3.8e-67;
Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

OY 1 AEIYKDKGKVDLYGKAVGLHFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 60
DB 15 AEIYKDKGKVDLYGKAVGLHFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 69
OY 61 WEYNFOGNSSEGADQGTGKTRLAFLAGLYADVSFDRGNGVVDALGYTDMLEPFG 120
DB 70 WESEFSGNKTESDSQ---KTRLAFLAGLYADVSFDRGNGVVDALGYTDMLEPFG 126
OY 121 D-TAYSDFEVRGVATYRNSNFEGLVDGLNFAVOYLGNEDTRARRS-NGDVG 179
DB 127 DSSAQTNMTKTRASGLAYRNTDFGLVDGLTQYOGKNGEAKQNGDVG 186
OY 180 YEEG--FGIYAGVADRTNLOEAQPLGNGKAEQMATGLKYDANNITYLAANGETR 237
DB 187 YDFGSGDFAVSAAYTSSDRTNDONLARQGSRAEMATGLKYDANNITYLAANGETR 246
OY 238 TPITNKFTNTSGFANKQDVLVAOYQDFGLRPSIATYKSKAKVEGI-GEVDLVN 297
DB 247 TPI-----SGGFANKQDVLVAOYQDFGLRPSIATYKSKAKVEGI-GEVDLVN 300
OY 298 GATYFNKMNSTVYDIINOIDSQKLGVSDD 330
DB 301 GLTYFNKMNSTVYDIINOIDSQKLGVSDD 333

RESULT 10
OYR59 PRELIMINARY: PRT: 333 AA.
AC OYR59: 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAY-2000 (TREMREL. 16, Last annotation update)
DE OUTER MEMBRANE PHOSPHATE PORIN PRECURSOR (FRAGMENT).
GN PHOE.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC9633T;
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,
Kemp D.J.;
RT "Phylogenetic evidence for reclassification of Calymmatobacterium
granulomatis as Klebsiella granulomatis comb. nov.";
RL Int. J. Syst. Bacteriol. 0:0-0(1999).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
(BY SIMILARITY).
CC EMBL: AF064793; AAD11630.1; -.
DR HSSP: P02932; 1PHO.
DR InterPro: IPR001702; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLIPEPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT NON_TER 1
FT NON_TER 333
SQ SEQUENCE 333 AA: 36855 MW: 68AF32C5077705EA CRC64;

Query Match 56.4%; Score 1019.5; DB 2; Length 333;
Best Local Similarity 58.9%; Pred. No. 5.3e-67;
Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

OY 1 AEIYKDKGKVDLYGKAVGLHFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 60
DB 15 AEIYKDKGKVDLYGKAVGLHFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 69
OY 61 WEYNFOGNSSEGADQGTGKTRLAFLAGLYADVSFDRGNGVVDALGYTDMLEPFG 120
DB 70 WESEFSGNKTESDSQ---KTRLAFLAGLYADVSFDRGNGVVDALGYTDMLEPFG 126

Qy	121	D-1A	SDDEFEVGRVGYAATRHNSFEPLDNGLFAVOYLCKNERDPAHNSNGCVGGS	179
	127	SSAQTDDFMKRRASGATLATRNTEDFELVYGLDITLQYCKNSGRKKNKGCVGSIS	186	
Qy	180	YEYEG--FGIAYGAGADRTNLQEAQPLGNGKRAEQWATGLKTDANNIYLAANYGERNA	237	
Db	187	YDEGSSDPFAVAATSSDRTNDQVLLLRGSGSKAEAMATGLKTDANNIYLATMYSERRK	246	
Qy	238	TPITNKFNTISGFAKNTQDVLVLAQYQPDFGLRPSIAIYYSKAKADVEGIDVDLVNTEFV	297	
Db	247	TPITNKFNTISGFAKNTQDVLVLAQYQPDFGLRPSIAIYYSKAKADVEGIDVDLVNTEFV	300	
Qy	298	GATYVENKMSYVDYIIINQDSNKLKGVGSSD	330	
Db	301	GLTYYFNKNMNAFVDYKTNQLKSDNKLKGVGSSD	333	

RESULT	11		
Q9F889			
ID	Q9F889	PRELIMINARY;	PRT; 360 AA.
AC	Q9F889;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, last annotation update)		
DE	OUTER MEMBRANE PROTEIN F PRECURSOR (FRAGMENT).		
GN	OMPf.		
OS	Salmonella typhi.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
CC	Salmonella.		
OX	NCBI_TaxID=601;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=9, 12, (VI);D;		
RA	Mate-Gonzalez M.T., Pelayo R., Islas A., Montolio L.,		
RA	Ortiz-Navarrete V.F.;		
RT	"Ompf porin is expressed in Salmonella typhi.";		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBD databases.		
DR	EMBL; AF251685; AAC09474.1; "		
KW	Signal.		
FT	SIGNAL.	1 22	POTENTIAL.
FT	NON_TER	360	360
SQ	SEQUENCE	360 AA; 39665 MW; B48A51685965771 CRC64;	

Query Match	56.28;	Score 1017.5;	Length 360;
Best Local Similarity	57.68;	Pred. No. 8.3e-67;	
Matches 198; Conservative	54;	Mismatches 79;	Indels 13; Gaps 8;

```

Qy 1 AIIYKDKNKVDLYKAAVGLHYFSGKNEEMNYSNGCMGTARLFCKEPTIINSLDYGVC 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 23 AETIYKDKNKRLDYKAVGRHWMTT -TODSK ---NAOYTAQISGKEPTIINTDLGFPQ 78

Qy 61 WEYNFOGNNSEGADQOTGNKTRLPACLTADVGSFDTGRNYGVYDALTITMLPEFGC 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 79 WEYRTKAPRAAE -EBOONSLVRLAFAGLTAEVSDIDYGRNYGLVYDESYTMAPFFSG 137
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 121 DE - - -AYSDDPFVGVGVATYVRNNSNPEGLVDGINFAVOYTGKNERDPTARRS -NGDGVGC 176
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 ETWGGATYVDNMYTNSAGGLLTRYNSDFEGLVDGLSFGIOYQGNKQDHSIINQNGGVGC 197
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 177 S1SYSEGFIVGATGAADRTNLOEAQPLGNGKKAEDOMATISGLKTDANNITLYANTGETRN 236
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198 TMAVEDEGGVYAAASNSKRTINDODDR -GNGDRAESNAVAGAKDANNVLYLAVYATRN 256
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 237 AMPITNKFFNTSGFANRQDVLVLAOYQFDGLRPISLARYTSKAKADEGI -GDVLTNYF 295
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 257 MS1VEVETVTDYEMANKQNTQNELVAAQYQFDGRLPALSYVSGSKGQKOLNAGDSADLAKYI 316

Qy 296 EVGATYFENKKNSTVYDITINOIDSNDKLG--VGSDDTVANGVI 337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 317 QAGATYFENKKNMNVWDYRENLLEBNDSSSYSGTIDDOAAAVGI 360

```

RESULT	12			
SEQUENCE	087753	PRELIMINARY:	PRT:	359 AA.
ID	087753			
AC	087753;			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	OMP35 PORIN PRECURSOR.			
GN	OMP35.			
OS	Klebsiella pneumoniae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
CC	Klebsiella.			
OX	NCBI_TaxID=573;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KT755;			
RA	Hernandez-Alles S.;			
RL	Thesis (1998), Universitat de les Illes Balears, Valencia, Spain.			
DR	EMBL: AJ011501; CAA09665.1; -.			
DR	HSSP; P02931; IGFN.			
DR	InterPro: IPR001702; -.			
DR	Pfam: PF00267; Gram-var_porins; 1.			
DR	PRINTS; PR00182; ECOLIETPORIN.			
KW	Signal.			
FT	SIGNAL.	1	22	POTENTIAL.
FT	CHAIN	23	359	OMP35 PORIN.
SQ	SEQUENCE	359 AA;	39510 MW;	12D516340058E58B CRC64;

Query Match	51.7%	Score 934.5;	DB 2;	Length 359;
Best Local Similarity	55.3%	Pred. No. 9.6e-61;		
Matches 199; Conservative	42;	Mismatches 90;		Gaps 11;

QY	1	AAIYKDGKRDVLYKAKVGLHYFSKNGCNEENSYGNGDMTARGLGKEPQINSDLTYGQ	60
QY <td>23</td> <td>AAIYKNGKNGKIDFYKGMKGEHWMT--NDDTS---SDTPYAIRGLKKEETIINOQLGYGQ</td> <td>78</td>	23	AAIYKNGKNGKIDFYKGMKGEHWMT--NDDTS---SDTPYAIRGLKKEETIINOQLGYGQ	78
QY	61	WEYNGNGNSGADQOQONTKRLAFAGLKYADVGSFQDGRNYGVVYALQYDTMLPEFGG	120
Db	79	WEYNNDASHWES--SQT-TKTRLAFAGLKAAGEYGSFDYGRNYGAIIVYEATDMLVPMWG	135
QY	121	D-TAYSDDPEFVYRGVATYTRNSNFFGLVDGLNFAVQYLAKNEBDTA-RSNSDGVGGS	178
Db	136	DGMNTDNTMTRTGNGVATYTRNSDPEGLVDLSFALQYQGNHDBRAIRQNGDGFSTAA	195
QY	179	SYEYE-GEFGIVAGYGAADRTNLQEAQPLGNGKKAQOMATGLKYANNIYLAANYGETRNA	237
Db	196	TYAFDGIATLSAGYSSNRSYQDKAD--GNGDKAEAMATSAKYDANNIYAAMVYSQTYNM	253
QY	238	TPITNKTFTSGFANKTDQVLLVAQYQDFGLRPSIATYKSKADVEGI-----GDVDLV	292
Db	254	TP-----EEDNHFACKTONFEAVVQYQDFGLRPSIGVQYQKGLDLSRAGFSGGADLV	308
QY	293	NFEVGCATYFENKNNSTYVDYLIINOJDEDN--KLGVSDDTVAVGYIYQF	340
Db	309	KYLEVGATWYFENKNNMVAIAKFNOLDNDNDDTKAAGVATDQAAVGIYQF	359
RESULT	13		
Q9EXH8			
ID	Q9EXH8	PRELIMINARY:	PRT: 315 AA.
AC	Q9EXH8		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	OMPK35 PROTEIN PRECURSOR.		
GN	OMPK35.		
OS	Klebsiella pneumoniae.		
OC	Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Klebsiella.		
OX	NCBI_TaxID=573;		
RN	[1]		
RP	SEQUENCE FROM N.A.		


```

OY 233 GYGLSGSGTSGRGLGCGGAGAAAA-----AAAAAGAGCGG 271
DB 326 GYGG19SGG-agTGG199GGaGaaAaaagaggg19sggagggagaaAaaagagggg 384
OY 272 YGGLSGG 279
DB 385 YGG19SGG 392

RESULT 3
AAW27178
ID AAW27178 standard; Protein: 646 AA.
XX
AC AAW27178;
XX
DT 09-DEC-1997 (first entry)
XX
DE Nephila clavipes spider silk protein.
XX
KM High strength film; fibre; woven article; parachutes; sails;
KW absorber; body armour; heavy metal; biological weapon; chemical;
XX flavour; fragrance; Nephila clavipes.
OS Nephila clavipes.
XX
PN MO9708315-A1.
XX
PD 06-MAR-1997.
XX
PF 22-AUG-1996; 96MO-US13767.
XX
PR 22-AUG-1995; 95US-0517694.
XX
PA (BASE/) BASEL R. M.
PA (ELIO/) ELION G. R.
XX
PI Basel RM, Elion GR;
XX
DR WPI; 1997-179272/16.
XX
N-PSDB; AAT85356.
XX
PT New opt. multimerised DNA sequences encoding spider silk protein -
XX making high strength films, fibres, woven articles etc.
XX
XX Example 2; Fig 1; 57pp; English.
XX
PS A process has been developed for the production of a DNA fragment
XX encoding silk protein. The process involves: (a) selecting target DNA,
XX from a silk-producing spider, that contains many repetitive and non-
XX repetitive regions; (b) selecting a single-stranded DNA primer of at
XX least 10 nucleotides with a sequence that is complementary to a region
XX of the target; (c) repetitively combining the primer with melted target
XX DNA, incubating the mixture with nucleotides and a DNA polymerase with
XX proofreading activity to produce a DNA fragment which is complementary
XX to the target and is at least 2 kb long. The present sequence encodes
XX the spider silk protein from Nephila clavipes. The DNA fragment can be
XX used to make fibres, films, woven articles, e.g. for use in parachutes,
XX sails, body armour, and absorbers (e.g. of heavy metals, biological
XX weapons, DNA, chemicals, flavours and fragrances). The high molecular
XX weight (90-250 kD) of spider silk proteins can be produced on a
XX commercial scale (at over 2 g/1 cell mass). It has better tensile
XX strength and elasticity than silkworm silk. Inclusion of both repetitive
XX and non-repetitive regions ensures isolation of stable clones.
XX
SQ Sequence 646 AA:

```

```

Query Match 69.4%; Score 1031; DB 18; Length 646;
Best Local Similarity 55.5%; Pred. No. 8.4e-73;
Matches 244; Conservative 6; Mismatches 10; Indels 180; Gaps 13:
OY 11 GSMASGRGLGCGGAGAAAA-----AAAAAGAGCGGGLG 49

```

```

DB 50 GSGGagTggl199GgaGaaAaaagaggg19sggagggagaaAaaagagggg19gs 109
OY 50 QGT-----SGRGLGCGGAGAAAA-- 68
DB 110 qgaagTggaAaaagaggggYggaaggggYggl9sggagTggl199GgaGaaAaa 169
OY 69 -----AAAAAGAGCGGCGYGLGSGTSGRGLGCGGAGAAAA 109
DB 170 aggaaggg199GgaGaaAaaagaggggYggl9sggagTggl199GgaGaaAaa 224
OY 110 AAAGAGCGGCGYGLGSGTSGRGLGCGGAGAAAAAG-----GAGCGG 159
DB 225 aaagaggggYggl9sgg-agTggl199GgaAaaAaaagagggYggl199GgaGaa 283
OY 160 GLSGSGTSGRGLGCGGAGAAAA----- 182
DB 284 g19sgg-agTggl199GgaAaaAaaagaggg19sggagTggl199GgaAaaAaa 342
OY 183 -----AAAAAGAGCGGCGYGLGSGTSGRGLGCGGAGAAAA-- 225
DB 343 199GgaTggaAaaAaaagaggggYggl9sgg-agTggl199GgaAaaAaaagag 401
OY 226 -----AGAGCGGCGYGLGSGTSGRGLGCGGAGAAAA-----A 259
DB 402 gYggl199GgaGaaAaaagagggYggl9sgg-sgTggl199GgaAaaAaaagag 460
OY 260 AAAAGAGAGCGGCGYGLGSGG 279
DB 461 aaaaagaggggYggl9sgg 480

RESULT 4
AA40097
ID AA40097 standard; protein; 651 AA.
XX
AC AA40097;
XX
DT 19-NOV-1999 (first entry)
XX
DE Spider silk protein spiderine major 1.
XX
KM Spider silk protein; spiderine major 1; cosmetic; make-up;
KW dermatological compositions; hair care; skin care; sunscreen;
XX hormone; moisturizer; skin disorder; skin disorder.
OS Nephila clavipes.
XX
PN FR2774588-A1.
XX
PD 13-AUG-1999.
XX
PF 11-FEB-1998; 98FR-0001614.
XX
PR 11-FEB-1998; 98FR-0001614.
XX
PA (OREA ) L'OREAL SA.
XX
PI Philippe M, Garson JC, Arraudeau JP;
XX
DR WPI; 1999-510729/43.
XX
PT Cosmetic or dermatological composition containing spider silk protein,
XX for hair or skin care, in make-up or sunscreens
XX
PS Claim 3; Fig 1; 32pp; French.
XX
XX The present sequence represents the natural spider silk protein
XX spiderine major 1. The protein improves the moisturizing/softening
XX action of the compositions. The protein, and its fragments are used
XX in cosmetic or dermatological compositions. These compositions have
XX use as hair or skin care products; and make-up or sunscreens.
XX As the protein is a good, persistent film-formers on the skin

```

CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.

XX Sequence 651 AA;

Query Match 68.4%; Score 1015; DB 20; Length 651;
Best Local Similarity 59.2%; Pred. No. 1.5e-71;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

```
QY 11 GSMASGRGGLGCGGAGAAAA-----AAAAAGAGCGCGYG 45
  ||:|||||
Db 93 gsgqagrtg1g9qgagaaaaaagagaggyg1gngqagrtg9qgaaaaaagagaggyg 152
  ||:|||||
QY 46 GLGSGGTGSRGGLGCGGAGAAAA-----AGAGCGGTGGLGSGGTGSRGGL 96
  ||:|||||
Db 153 g1gsgg-agrtg1g9qgagaaaaaagagaggyg1g9qgagaggyg1gsgg-agrtg1 210
  ||:|||||
QY 97 GGCGAGAAAA-----AAAAAGAGCGGTGGLGSGGTGSRGGLG 135
  ||:|||||
Db 211 g9qgagaaaaaagagaggyg1g9qgagagagaaaaaagagaggyg1gsgg----agrtg 266
  ||:|||||
QY 136 GCGAGAAAA-----AGAGCGGTGGLGSGGTGSRGGLGCGGAGAAAA----- 182
  ||:|||||
Db 267 gsgagaaaaaagagaggyg1g9qgagaggyg1gsgg-agrtg1g9qgagaaaaagag 325
  ||:|||||
QY 183 -----AAAAAGAGCGGTGGLGSGGTGSRGGLGCGGAGAAAA----- 220
  ||:|||||
Db 336 g9g1g9qgagagagagaggyg1gsgg-agrtg1g9qgagagaaaaaagagag 384
  ||:|||||
QY 221 -----AAAAAGAGCGGTGGLGSGGTGSRGGLGCGGAGAAAA-- 259
  ||:|||||
Db 365 g9g1g9qgagagagagagagagaggyg1gngq-agrtg1g9qgagaaaaaa 443
  ||:|||||
QY 260 -----AAAAAGAGCGGTGGLGSGGTGSRGGLGSGGTGSRGGLG 279
  ||:|||||
Db 444 g9agaggyg1gngqagrtg9qgaaaaaagagaggyg1gsgg 485
```

RESULT 5

AAW53346 standard; Protein: 718 AA.

XX AC AAW53346;

XX DT 06-JUL-1998 (first entry)

XX DE Nephila clavipes spider silk protein.

XX KW Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;

XX KM cocoon; tensile strength; elasticity.

XX OS Nephila clavipes.

XX PN US5728810-A.

XX PD 17-MAR-1998.

XX PF 19-APR-1995; 95US-0425069.

XX PR 15-APR-1991; 91US-0684819.

XX PR 20-APR-1990; 90US-0511792.

XX PR 04-OCT-1994; 94US-0317844.

XX PR 19-APR-1995; 95US-0425069.

XX PA (UYWY-) UNIV WYOMING.

XX PT Himan MB, Lewis RV, Xu M;

XX WPI; 1998-270437/24.

XX DR N-PSDB; AAV23249.

XX PT Recombinant spider silk proteins - useful for making fibres
XX PS Claim 1; Column 29-34; 68pp; English.

CC The present sequence represents a spider silk protein from the present
CC invention. Spider silk proteins, and peptide fragments within the
CC proteins, can be produced and purified independently and can then be
CC mixed and made into fibres that have higher tensile strengths and
CC elasticity than naturally occurring fibres. The fibres can be used in
CC mixed composites. The invention allows the two naturally occurring
CC Nephila clavipes silk proteins to form silk fibres of high tensile strength
CC and elasticity.

XX Sequence 718 AA;

Query Match 68.4%; Score 1015; DB 19; Length 718;
Best Local Similarity 59.2%; Pred. No. 1.6e-71;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

```
QY 11 GSMASGRGGLGCGGAGAAAA-----AAAAAGAGCGCGYG 45
  ||:|||||
Db 93 gsgqagrtg1g9qgagaaaaaagagaggyg1gngqagrtg9qgaaaaaagagaggyg 152
  ||:|||||
QY 46 GLGSGGTGSRGGLGCGGAGAAAA-----AGAGCGGTGGLGSGGTGSRGGL 96
  ||:|||||
Db 153 g1gsgg-agrtg1g9qgagaaaaaagagaggyg1g9qgagaggyg1gsgg-agrtg1 210
  ||:|||||
QY 97 GGCGAGAAAA-----AAAAAGAGCGGTGGLGSGGTGSRGGLG 135
  ||:|||||
Db 211 g9qgagaaaaaagagaggyg1g9qgagagagaaaaaagagaggyg1gsgg----agrtg 266
  ||:|||||
QY 136 GCGAGAAAA-----AGAGCGGTGGLGSGGTGSRGGLGCGGAGAAAA----- 182
  ||:|||||
Db 267 gsgagaaaaaagagaggyg1g9qgagaggyg1gsgg-agrtg1g9qgagaaaaagag 325
  ||:|||||
QY 183 -----AAAAAGAGCGGTGGLGSGGTGSRGGLGCGGAGAAAA----- 220
  ||:|||||
Db 336 g9g1g9qgagagagagaggyg1gsgg-agrtg1g9qgagagaaaaaagagag 384
  ||:|||||
QY 221 -----AAAAAGAGCGGTGGLGSGGTGSRGGLGCGGAGAAAA-- 259
  ||:|||||
Db 385 g9g1g9qgagagagagagagaggyg1gngq-agrtg1g9qgagaaaaaa 443
  ||:|||||
QY 260 -----AAAAAGAGCGGTGGLGSGGTGSRGGLGSGGTGSRGGLG 279
  ||:|||||
Db 444 g9agaggyg1gngqagrtg9qgaaaaaagagaggyg1gsgg 485
```

RESULT 6

AAV59070 standard; Protein: 718 AA.

XX AC AAV59070;

XX DT 08-MAR-2000 (first entry)

XX DE N. clavipes spider silk protein 1.

XX KW Spider silk protein; dragline silk protein; major ampullate gland; fiber.

XX OS Nephila clavipes.

XX PN US5989894-A.

XX PD 23-NOV-1999.

XX PF 04-OCT-1994; 94US-0317844.

XX PR 15-APR-1991; 91US-0684819.

XX PR 20-APR-1990; 90US-0511792.

XX (UYWY-) UNIV WYOMING.
 XX
 XX Human MB, Xu M, Lewis RV;
 XX
 XX MPI: 2000-061225/05.
 XX N-PSDB: AA238195.
 XX
 XX Isolated DNA, vector and transformed cell encoding for and useful in
 XX the production of spider silk protein -
 XX
 XX Claim 1; Fig 6A-D; 65pp: English.
 XX
 XX The invention provides isolated cDNA molecules coding for spider silk
 XX proteins. The spider silk proteins are characterized by repeating alpha
 XX and beta regions and optional variable regions. The DNA sequences are
 XX useful in the production of spider silk protein by recombinant DNA
 XX techniques. The recombinant spider silk proteins may be used for the
 XX production of fibers. The present sequence represents the spider silk
 XX protein 1, derived from the major ampullate gland of Nephila clavipes.
 XX
 XX Sequence 718 AA;
 XX
 XX Query Match 68.4%; Score 1015; DB 21; Length 718;
 XX Best Local Similarity 59.2%; Pred. No. 1.6e-71;
 XX Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;
 XX
 XX 11 GSMASGRGCLGCGGCAAAA-----AAAAAGAGCGCGC 45
 XX || :|||||
 XX 93 gsggagrglggggagaaaaaaggaggygglgngagatrgggaaaaaaggaggyg 152
 XX
 XX 46 GLSGSGTGRGGLGCGGCAAAA-----AGAGCGGCGTGLSGTSGRGL 96
 XX ||||| :|||||
 XX 153 glsgsg-egrglggggagaaaaaaggaggygglgngagaggygglgsgg-egrggl 210
 XX
 XX 97 GCGGCAAAAA-----AAAGGAGCGGCGTGLSGTSGRGLG 135
 XX |||||
 XX 211 ggggagaaaaaaggaggyglggygagggagaaaaaggaggygglgsgg----agrg 266
 XX
 XX 136 GCGGCAAAAA-----AGAGCGGCGTGLSGTSGRGLGCGGCAAAA----- 182
 XX ||:|||||
 XX 267 gegagaaaaaggaggygglgngagaggygglgsgg-egrgglgngagaaaaaggag 325
 XX
 XX 183 -----AAAAAGGAGCGGCGTGLSGTSGRGLGCGGCAAAA----- 220
 XX |||||
 XX 326 gggglgagagagaaaaaaggaggygglgsgg-egrgglgngagavaaaaaggagag 384
 XX
 XX 221 -----AAAAAGGAGCGGCGTGLSGTSGRGLGCGGCAAAA-- 259
 XX |||||
 XX 385 ggygglgagaggygagaaaaaaggaggygglgngg-egrgglgngagaaaaaa 443
 XX 260 -----AAAGGAGCGGCGTGLSGG 279
 XX |||||
 XX 444 gggaggygglgngagagrgggagaaaaaggaggygglgsgg 485
 XX
 XX RESULT 7
 XX AARI4308
 XX ID AARI4308 standard; Protein: 718 AA.
 XX
 XX AARI4308;
 XX
 XX 15-JAN-1992 (first entry)
 XX
 XX N.clavipes dragline silk protein-1.
 XX
 XX DE protein superfibre; major ampullate silk; orb web spider.
 XX
 XX Nephila clavipes.
 XX
 XX EP452925-A.

```
PD      23-OCT-1991.
XX
XX PF      18-APR-1991;       91EP-0106217.
XX
XX PR      20-APR-1990;       90US-0511792.
XX
XX (UYWY-) UNIV OF WYOMING.
PA
XX Lewis RV, Xu M, Hlman M;
PI
XX MPI: 1991-31219/43.
DR      N-PSDB: AAO1A4I83.
XX
XX DNA encoding spider silk protein-1 and 2 and variants - isolated
PT from Nephila clavigipes, for prodn. of spider silk protein and
PR fibres having desired characteristics
XX
XX Claim 15; Page 23; 48pp; English.
PS
XX
XX The spider silk protein contains a basic 34 amino acid repeat. The
CC repeat itself contains 3 regions. The first comprises 0-9 amino
CC acids with a sequence AGR(GGX)2. This region is not highly
CC conserved. The second region has a sequence GAG(AIX) which is highly
CC conserved and is 8-10 amino acids long. The third segment is (GGX)5
CC and is 15 amino acids long and is very highly conserved. In most
CC cases X is A,Q, Y or L. Removal of the poly-(Ala) segments results
CC in a silk having lower elasticity.
XX
XX Sequence      718 AA;
SQ
Query Match          67.4%; Score 1001; DB 12; Length 718;
Best Local Similarity 58.7%; Pred. No. 1.9e-70;
Matches 236; Conservative 8; Mismatches 16; Indels 142; Gaps 13;
OY      11 GSMASGRGGLSGOGAGAANA-----AAAAACGACGCGTG 45
    II :|||||
DB      93 gsgagagrgl1gggagaagaaaggaggyggl1gnqgagr1gggaagaaaggaggyg 152
OY      46 GLGSQGTSGRGGLSGOGAGAAAANA-----AGCAGCGGYGGLSGQTSGRGSL 96
    I|||
DB      153 gl1esq1aaigrgl1gggagaaaaaaggaggyggl1gggagaggyggl1gsqg1-agrgrl 210
OY      97 GCGAGAAAANA-----AAAGAGCGGYGGLSGQTSGRGGLG 135
    |||||
DB      211 gsgagagaaaaaaggaggygl1gggagaggaagaaaaggaggyggl1gsqg----agrg 266
OY      136 GCGAGAAAANA-----AGCAGCGGYGGLSGQTSGRGGLGCGCAANA--- 182
    I:|||||
DB      267 gsgagaaaaaaggaggyggygl1gggagaggyggygl1gsqg1-agrgrgl1gggagaaaaaggag 325
OY      183 -----AAAAAGCAGCGGYGGLSGQTSGRGGLGCGCAANA----- 220
    |||||
DB      336 ggg11gggagaggaagaaaaaaggaggyggl1gsqg1-agrggl1gggagavaaaaggagdg 384
OY      221 -----AAAAAGCAGCGGYGGLSGQTSGRGGLGCGAGAAAAN-- 259
    |||||
DB      385 ggyggl1gsgagagrggagagaaaaaaggaggyggl1gmng1-agrggl1gggagaaaaaaa 443
OY      260 -----AAAAAGCAGCGGYGGLSGQG 279
DB      444 ggaaggyggygl1gnqgagrggyggaagaaaggaggyggl1gsqg 485
RESULT      8
ID      AAR9057
XX      AAR9057 standard; Protein; 604 AA.
XX
XX AAR9057;
XX
XX 17-JAN-1997 (first entry)
XX
XX Spider dragline variant. DE-1B.16 polymer.
```

XX Spider: dragline protein; variant; monomer; polymer;
 KW fibre forming region; Spidroin 1; Nephrila clavipes; DP1; mimic;
 KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KW rope; surgical suture; implant; reinforcement; film; coating.
 XX Synthetic.
 OS
 PN WO9429450-A2.
 PD 22-DEC-1994.
 XX 15-JUN-1994; 94WO-US06689.
 PF 15-JUN-1993; 93US-0077600.
 PR 15-JUN-1993; 93US-0077600.
 XX (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA Fahnestock SR;
 PI
 P1
 XX
 XX
 DR WPI: 1995-036479/05.
 XX
 PT New synthetic variants of spider dragline protein - for making
 PT fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc., also related DNA, vectors and transformed cells
 PS
 XX Disclosure; Page 106-108; 168pp; English.
 XX This sequence represents a synthetic spider dragline variant polymer,
 CC DP-1B.16. The sequence of the DP-1B.16 monomer is given in AAR99056.
 CC The polypeptide monomer is a variant based on a consensus sequence
 CC derived from the fibre forming regions of spider dragline protein,
 CC esp. the natural protein 1 (Spidroin 1) from Nephrila clavipes. The
 CC DNA sequence encoding the monomer may be used in the recombinant
 CC production of the variant protein in a recombinant host, e.g. E. coli
 CC or Bacillus subtilis. Synthetic analogues of DP1 were designed to mimic
 CC the repeating consensus sequence of the natural protein and the pattern
 CC of variation among individual repeats. This monomer exhibits all of the
 CC regularities of (1)-(5) below. In addition, it exhibits a regularity of
 CC the natural sequence which is not shared by DP-1A, namely that a repeat
 CC in which both GYG and GRG are deleted is generally preceded by a repeat
 CC lacking the entire poly-alanine repeat, with one intervening repeat.
 CC The sequence of DP-1B matches the natural sequence more closely over
 CC a more extended segment than does DP-1A. The individual repeats differ
 CC from the consensus sequence given in AAW06201 according to the pattern:
 CC (1) the poly-alanine sequence varies in length from 0-7
 CC residues; (2) when the entire poly-alanine sequence is deleted,
 CC so also is the surrounding sequence encompassing AGRGIGCGAGAACG;
 CC (3) aside from the poly-alanine sequence, deletions usually
 CC encompass integral multiples of three consecutive residues;
 CC (4) deletion of GYG is generally accompanied by deletion of GRG
 CC in the same sequence; and
 CC (5) a repeat in which the entire poly-alanine sequence is
 CC deleted is generally preceded by a repeat containing six alanine
 CC residues.
 CC The proteins may be used to produce fibres of high tensile strength and
 CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
 CC for implants, plastic reinforcements, films, coatings, etc.
 XX
 XX Sequence 604 AA;
 XX
 Query Match 66.3%; Score 984; DB 16; Length 604;
 Best Local Similarity 60.2%; Pred. No. 3.4e-69;
 Matches 231; Conservative 4; Mismatches 13; Indels 136; Gaps 13;
 OY 11 GSMASGRGLGCGAGAAA-----AAAAAGAGCGGCGGCGS 49
 DB 13 gsgaggrgglgsgagagaaagagagggglsqgagqagagaaagaggggylgs 72
 OY 50 OCTSRGGIGGGGAGAAAAAGAGGCGGCGGCGT-----SGRGGCG 97
 DB 73 qg-----agrggagag-----aaaaagagaggggylgsqagqggylgsqaggrgglg 124

OY 98 GCGAGAAA-----AAAAAGAGCGGCGGCGTSGRGCGIG 136
 DB 125 gsgagagagagagagggglgsgagagqagagaaagaggggylgsqg-----agrgg 180
 OY 137 OGAGAAAAAGAGAGCGGCGGCGTSGRGCGCGAGAGAAA-- 182
 DB 181 qgag-----aaaaagagaggggylgsqagaggggylgsqaggrgglgsqagagaa 236
 OY 183 -----AAAAAGAGCGGCGGCGTSGRGCGAGAGAAAA 223
 DB 237 aggagggglgsqagagagagaaagaggggylgsqg-----agrggagagagaaag 292
 OY 224 AAAG-----GAGCGGCGGCGTSGRGCGAGAGAAA----- 258
 DB 293 gagggggylgsqagaggggylgsqg-agrgglgsqagagagagagaggggylgsqagag 351
 OY 259 ---AAAAAGAGCGGCGGCGTSGRGCGAGAGAAA----- 375
 DB 352 gagagagagagaggggylgsqg 375
 RESULT 9
 AAR99055
 ID AAR99055 standard; Protein: 606 AA.
 XX
 AC AAR99055;
 XX
 DT 17-JAN-1997 (first entry)
 XX
 DE Spider dragline variant, DP-1B.9 polymer.
 XX
 KW Spider: dragline protein; variant; monomer; polymer;
 KW fibre forming region; Spidroin 1; Nephrila clavipes; DP1; mimic;
 KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KW rope; surgical suture; implant; reinforcement; film; coating.
 XX
 OS Synthetic.
 PN WO9429450-A2.
 PD 22-DEC-1994.
 XX 15-JUN-1994; 94WO-US06689.
 PF 15-JUN-1993; 93US-0077600.
 PR 15-JUN-1993; 93US-0077600.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX Fahnestock SR;
 PI
 P1
 XX
 DR WPI: 1995-036479/05.
 XX
 PT New synthetic variants of spider dragline protein - for making
 PT fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc., also related DNA, vectors and transformed cells
 PS
 XX Disclosure; Page 88-90; 168pp; English.
 XX This sequence represents a synthetic spider dragline variant polymer,
 CC DP-1B.9. The sequence of the DP-1B.9 monomer is given in AAR99054.
 CC The polypeptide monomer is a variant based on a consensus sequence
 CC derived from the fibre forming regions of spider dragline protein,
 CC esp. the natural protein 1 (Spidroin 1) from Nephrila clavipes. The
 CC DNA sequence encoding the monomer may be used in the recombinant
 CC production of the variant protein in a recombinant host, e.g. E. coli
 CC or Bacillus subtilis. Synthetic analogues of DP1 were designed to mimic
 CC the repeating consensus sequence of the natural protein and the pattern
 CC of variation among individual repeats. This monomer exhibits all of the
 CC regularities of (1)-(5) below. In addition, it exhibits a regularity of
 CC the natural sequence which is not shared by DP-1A, namely that a repeat
 CC in which both GYG and GRG are deleted is generally preceded by a repeat
 CC lacking the entire poly-alanine repeat, with one intervening repeat.

CC The sequence of DP-1B matches the natural sequence more closely over
 CC a more extended segment than does DP-1A. The individual repeats differ
 CC from the consensus sequence given in AM06501 according to the pattern:
 CC (1) the poly-alanine sequence varies in length from 0-7
 CC residues; (2) when the entire poly-alanine sequence is deleted,
 CC so also is the surrounding sequence encompassing AGKGLGQAGANGG;
 CC (3) aside from the poly-alanine sequence, deletions usually
 CC encompass integral multiples of three consecutive residues;
 CC (4) deletion of GYG is generally accompanied by deletion of GRG
 CC in the same sequence; and
 CC (5) a repeat in which the entire poly-alanine sequence is
 CC deleted is generally preceded by a repeat containing six alanine
 CC residues.
 CC The proteins may be used to produce fibres of high tensile strength and
 CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
 CC for implants, plastic reinforcements, films, coatings, etc.
 XX Sequence 606 AA;
 XQ

Sequence 606 AA;

Query Match	65.78;	Score 976;	DB 16;	length 606;
Best Local Similarity	59.18;	Pred. NO. 1.4e-68;		
Matches 230; Conservative	4;	Mismatches 11;	Indels 144;	Gaps 13

OY	11	GSMAAGCGGCGGAGAAA-----AAAAAGAGCGGCGGAGS	4
Db	11		70
OY	11	gsqgaag-yggl-ygqgaagaaagagaggg-l-ygqgaagaaagaaagagaggg-l-yg	70
Db	11		70
OY	50	QGTSGRGGLGCGGAGAAAAAAGAGCGGCGGAGLSGQT-----SGRGGLG	97
Db	71		122
OY	98	GCGAGAAA-----AAAAAGAGCGGCGGAGLSGQTSGRGGLG	133
Db	123		171
OY	137	QGAGAAAAAAGAGAGCGGCGGAGLSGQT-----SGRGGLGCGGAGAAA--	180
Db	179		233
OY	183	-----AAAAAGAGCGGCGGAGLSGQTSGRGGLGCGGAGAAAAA	222
Db	235		280
OY	224	AAAGAGCGGCGGAGLSGQT-----SGRGGLGCGGAGAAA-----	253
Db	287		344
OY	259	AAAAAGAGAGCGGCGGAGLSGQT	279
Db	347		375

RESULT 10

ID AAY40101 standard; protein; 606 AA.

AC AAY40101;

DT 19-NOV-1999 (first entry)

DE Polymer of an analogue of spider silk protein spiderone major 1.

KW Spider silk protein; spideroïne major 1; cosmetic; make-up;
KW dermatological compositions; hair care; skin care; sunscreen
KW hormone; moisturizer; skin disorder; skin disorder.

OS Synthetic.

Nephila clavipes.

Key Peptide	Location/Qualifiers
FH	1.101
FT	
FT	

XX
PN FR2774588-A1

PD 13-AUG-1999

PF 11-FEB-1998; 98FR-0001614

PR 11-FEB-1998; 98FR-0001614

PA (OREA) L'OREAL SA

PI Philippe M, Garson JC, Arraudeau JP,

DR WPI; 1999-510729/43

PT Cosmetic or dermatological composition containing spider silk protein, PT for hair or skin care, in make-up or sunscreens -

PS Clatm 8; Flg 5B; 32pp; French.

The present sequence represents a polymer of an analogue of the spider silk protein spiderone major 1. The protein improves the moisturizing/softening action of the compositions. The protein, and its fragments are used in cosmetic or dermatological compositions. These compositions have use as hair or skin care products; and make-up or sunscreens. As the protein is a good, persistent film-formers on the skin of low surface density, it can be used for delivery of active agents that are generally difficult to administer, e.g. vitamins, hormones, moisturizers or agents for treating disorders of the skin and hair.

SQ Sequence 606 AA

Query Match	65.78;	Score 976;	DB 20;	Length 606;
Best Local Similarity	59.18;	Pred. No. 1.4e-68;		
Matches 230; Conservative	4;	Mismatches 11;	Indels 144;	Gaps 13

OY	11	GSMSGRCGLGCGAGAAA-----	AAAAAAAAAGAGCGTGTGCG	49
Db	13	gsgagagrygylgsgagagaaaaaagagagggylgsgagagagaaaaaagaggggylgsg		72
OY	50	QGTSGRCGLGCGAGAAA-----	AAAAAAAAAGAGCGTGTGCG	97
Db	73	gq-----agrgggagag-----	aaaaaagagagggylgsgagagggylgsgagagrgylg	124
OY	98	GCGAGAAA-----	AAAAAAAAAGAGCGTGTGCG	156
Db	125	gsgagaaaaaagagagggylgsgagagggagagaaaaaagagggylgsgq---	agrgg	180
OY	137	QCGAGAAA-----	AAAAAAAAAGAGCGTGTGCG	182
Db	181	qgag-----	aaaaaagagagggylgsgagagggylgsgagagrgylgsgagaaaaa	233
OY	183	-----	AAAAAAGAGCGTGTGCG	223
Db	237	agsgagggylgsgagagggagagaaaaaagagggylgsgq---	agrgggag-----	288
OY	224	AAAGAGCGTGTGCG	-----	258
Db	289	aaagagagggylgsgagagggylgsgagagrgylgsgagagagagagagagggylgsg		348
OY	259	-----	AAAAAAGAGCGTGTGCG	279
Db	349	gagagagagaaaaaagagagggylgsgq	377	

RESULT 11

ID AAY40102 standard; protein; 606 AA

AC MAY40102;

XX

```
DT 19-NOV-1999 (first entry)
XX
DE Polymer of an analogue of spider silk protein spidroine major 1.
KW Spider silk protein; spidroine major 1; cosmetic; make-up;
KM dermatological compositions; hair care; skin care; sunscreen;
KN hormone; moisturizer; skin disorder; skin disorder.
XX
OS Synthetic.
OS Nephtila clavipes.
XX
FH Key
FT Location/Qualifiers
FT Peptide
FT 1..101
FT /note: "monomer unit"
XX
PN FR2774588-A1.
XX
PD 13-AUG-1999.
XX
PE 11-FEB-1998; 98FR-0001614.
XX
PR 11-FEB-1998; 98FR-0001614.
XX
PA (OREA ) L'OREAL SA.
XX
PI Philippe M, Garson JC, Arraudau JP;
XX
DR WPI: 1999-510729/43.
XX
PT Cosmetic or dermatological composition containing spider silk protein,
PT for hair or skin care, in make-up or sunscreens
XX
PS Claim 8; Fig 6B; 32pp; French.
XX
XX The present sequence represents a polymer of an analogue of the spider
CC silk protein spidroine major 1. The protein improves the moisturizing/
CC softening action of the compositions. The protein, and its fragments
CC are used in cosmetic or dermatological compositions. These compositions
CC have use as hair or skin care products; and make-up or sunscreens.
CC As the protein is a good, persistent film-formers on the skin
CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.
CC
XX
SQ Sequence 606 AA:
Query Match 65.7%; Score 976; DB 20; Length 606;
Best Local Similarity 59.1%; Pred. No. 1.4e-68;
Matches 230; Conservative 4; Mismatches 11; Indels 144; Gaps 13;
OY 11 GSMASGRGLGOGAGAAAA-----AAAAAGAGGCGYGLG 49
DB 11 gsgagagrgl1gsgagagagagaggl1sggagagagagagagaggl1gs 70
OY 50 OGTSRGGLGOGAGAAAAAGAGGCGYGLGSGGT-----SGRGGLG 97
DB 71 gq-----agrggagag-----aaaaagagagggggl1sggagaggggl1sggagaggggl1sg 122
OY 98 GOGAGAAAA-----AAAAAGAGGCGYGLGSGGTSGRGGLG 136
DB 123 gggagagagagagagaggl1sggagagagagagagagagaggggl1sggagagagagagagagag 178
OY 137 OGAGAAAAAGAGAGGCGYGLGSGGT-----SGRGGLGOGAGAAAA-- 182
DB 179 gqag-----aaaaagagaggggl1sggagaggggl1sggagaggggl1sggagagagagagagag 234
OY 183 -----AAAAAGAGGCGYGLGSGGTSGRGGLGOGAGAAAAA 223
DB 235 aggagaggggl1sggagagagagagagagagaggggl1sggagagagagagagagagagagagagag 286
OY 224 AAAGAGGCGYGLGSGGT-----SGRGGLGOGAGAAAA----- 258
```

```
DB 287 aagagaggggl1sggagaggggl1sggagagrgl1gggagagagagagagaggggl1gs 346
OY 259 -----AAAAAGAGGCGYGLGSGGT 279
DB 347 ggaagagagagagagaggggl1sggag 375
RESULT 12
AAV40099
ID AAV40099 standard; protein: 615 AA.
XX
AC AAV40099;
XX
DT 19-NOV-1999 (first entry)
XX
DE Spider silk protein spidroine minor 1.
XX
KW Spider silk protein; spidroine minor 1; cosmetic; make-up;
KM dermatological compositions; hair care; skin care; sunscreen;
KN hormone; moisturizer; skin disorder; skin disorder.
XX
OS Nephtila clavipes.
XX
PN FR2774588-A1.
XX
PD 13-AUG-1999.
XX
PE 11-FEB-1998; 98FR-0001614.
XX
PR 11-FEB-1998; 98FR-0001614.
XX
PA (OREA ) L'OREAL SA.
XX
PI Philippe M, Garson JC, Arraudau JP;
XX
DR WPI: 1999-510729/43.
XX
PT Cosmetic or dermatological composition containing spider silk protein,
PT for hair or skin care, in make-up or sunscreens
XX
PS Claim 7; Fig 3; 32pp; French.
XX
XX The present sequence represents the natural spider silk protein
CC spidroine minor 1. The protein improves the moisturizing/softening
CC action of the compositions. The protein, and its fragments are used
CC in cosmetic or dermatological compositions. These compositions have
CC use as hair or skin care products; and make-up or sunscreens.
CC As the protein is a good, persistent film-formers on the skin
CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.
CC
XX
SQ Sequence 615 AA:
Query Match 45.6%; Score 677.5; DB 20; Length 615;
Best Local Similarity 47.7%; Pred. No. 1.7e-45;
Matches 167; Conservative 10; Mismatches 90; Indels 83; Gaps 13;
OY 11 GSMASGRGLGOGAGA-AAAAAGAGGCGYGT-----GLSGGTSGRGGLGQ--- 61
DB 60 gaggaggggggl1sggagagagagagagaggggl1sggagagagagagagagaggggl1gs 119
OY 62 --GAGAAAAAGAGAGGCGYGT-----GLSGGTSGR-----GLGOGAGAAAAA 109
DB 120 gaggagagagagagaggggl1sggagagagagagagagaggggl1sggagagagagagagagag 179
OY 110 AAAGAGGCGYGT-----GLSGGT-----SGRGGLGOGAGAAAAA 147
DB 180 agagagaggggl1sggagagagagagagagagagagagagagagagagagagagagagagagagagag 239
```

OY		1A8	AAAGCAGCGGCVGC-----LGSGCTGRCGLG---CGACAAAAAAAAGAAGCAG	192
Dd		240	AAGGAAGGGygygagagaataaaaggaggg-yrgagagaaaagaggaagsg	299
OY		193	QCGYGCLGSOGT-----SGRGGTGCGCAAAAAAAAAAGAGAQCGVGCLGSQG	241
Dd		300	Bgyyggygagagaaaaagagaggg-yryagagagaa--qagagyyggyggyg	357
OY		242	-----TSGRGCLGCGCACAAAAAAAAGAGAGCG--GYGCLGSQG	279
Dd		358	Agagagaaaaatgagggaygrgagagaaagtgtgagygyggyg	407
RESULT 13				
XX	ID	AAAR80168	standard; Protein; 831 AA.	
XX	XX	AAAR80168;		
XX	DT	01-MAY-1996	(first entry)	
DE		pMISST MASP spider silk protein insert product.		
KW		Spider silk; repeat unit; consensus; minor ampullate silk protein;		
XK		splidrocin; MASP; orb web spider; dragline.		
OS		Nephila clavipes.		
FH	Key	Location/Qualifiers		
FT	Misc-difference	272..285		
FT	/note=	"represented as indeterminate, corresponds to a highly compressed GC rich region in MAO98470 which could not be sequenced"		
PX		MO9525165-AI.		
PN		21-SEP-1995.		
PD		14-MAR-1995;	95WO-US03139.	
PE		14-MAR-1995;		
PR		14-MAR-1994;	94US-O209747.	
PA		(UYWY-) UNIT WYOMING.		
PX		Colein M, Lewis RV;		
PI		WPJ: 1995-336970/43.		
DR		N-PADB: MAO98470.		
XX		Polypeptide(s) comprising repeated unit amino acid sequences, also cDNAs - derived from minor ampullate spider silk proteins and used to form spider silk fibres		
PT		Claim 2; Fig 1; 8bp; English.		
PS	This sequence represents the product of cDNA clone, pMISst, encoding the orb web spider minor ampullate silk protein MISPL has been identified and sequenced. Repeat unit peptides of MISPL may form part of a larger polypeptide with an amino terminus (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit peptides make up spider silk proteins (splidrolins) which in turn aggregate to form the silk fibres. Spider silk fibres have high tensile strength and significant elasticity. An isolated cDNA clone of a silk protein encoding sequence is of use to produce the protein at high yields using recombinant DNA technology.			
Sequence	831 AA:			

Query Match	45.48;	Score 674;	DB 16;	Length 831;
Best Local Similarity	46.48;	Pred. No. 4.1e-45;		
Matches 168;	Conservative 7;	Mismatches 97;	Indels 90;	Gaps 13

```

OY      15  SGRGSLGGGAGAAAAAAGAGAGGCGGCG-----LGSQGTSGRGSLG--- 59
Db      313  agagagaggyrfgagagaaagagagaggyyggqsgyagagaaagagagaggyrfga 372
OY      60  GGGAGAAAAAAGAGAGAGCGCGYGLSSQGT-----SGRGSLGAGGAGAAAAA 108
Db      373  gagagagaaagagagaaagagaggyyggqsgyagagagaaagagagaggyrfgagagag 432
OY      109  AAAAGGAGCGGTGGLGSOG-----TSGRGSLGCGGAGAGAAAAAAGAGAGCG- 156
Db      433  aaa--gagagaggyggqsgyagagagaaatafgagagaggyrfgagagagaaagagagtg 490
OY      157  -GYGSLGSQGT-----SGRGSLG--GGGAGAAAAA-----AAAAGAGCGGCG 197
Db      491  agyggqsgyggagagagaaagagaggyrfgagagagaaagagagagagagaggyrfga 550
OY      198  GLGSQGT-----SGRGSLGCGGAGAAAAAAGAGAGCGGCGG 235
Db      551  ggggyrfgagagaaagagagaggyrfgagagagagagagagagaggyrfgqsgyrg 610
OY      236  ---GLGS-----GCTSGRGSLG--GGGAGAAAAAAGAGAGCGCGYGLSSQGTSGTRRP 286
Db      611  agagagaaagagagsgyggqsgyrgagagagaaagagagagaggyyggqsgyrgagagaa 670
OY      287  AA 288
Db      671  aa 672

```

RESULT 14

AAY28843
 ID AAY28843 standard; Protein; 641 AA.

AC AAY28843;

DT 17-JAN-2000 (first entry)

DE Epstein Barr Virus Nuclear Antigen 1 (EBNA 1).

Epstein Barr Virus Nuclear Antigen 1; plasmid PCMVENBNA; EBNA 1; episome; transfection; origin of replication; EBV oriP; receptor; eucaryotic host cell; recombinant cell line; ion channel; gene therapy multiple gene expression; transporter protein; transcription factor; adhesion molecule; antisense therapy; gene amplification; cell immortalisation.

OS Epstein-barr virus.

Key	Location/Qualifiers
FH	

FT	/note= "Corresponds to incomplete GG codon"
----	---

PN W09947647-A1

PD 23-SEP-1999.

PF 12-FEB-1999; 99WO-US03307.

PR 18-MAR-1998; 98US-0040961.

XX

XX

XX

DR N-PSDB; AAX90924.

XX

PT useful for g

PS Claim 2; Fig.

XX

New method for expressing genes from recombinant eukaryotic cells useful for gene therapy.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:47:50 ; Search time 48.42 Seconds
(without alignments)
121.068 Million cell updates/sec

Title: US-09-490-291-2

Perfect score: 1485
Sequence: 1 MRGSHHHHSGMASGRGL.....YGLGSGTSGIRPAKLN 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1015	68.4	718	1	US-08-425-069-2 Sequence 2, Appl1
2	1015	68.4	718	2	US-08-317-844B-2 Sequence 2, Appl1
3	1015	68.4	718	4	US-09-034-177-3 Sequence 3, Appl1
4	674	45.4	832	1	US-08-209-747-2 Sequence 2, Appl1
5	674	45.4	832	1	US-08-458-298-2 Sequence 2, Appl1
6	593.5	40.0	738	3	US-08-864-038A-3 Sequence 3, Appl1
7	561	37.8	235	2	US-08-529-190B-1 Sequence 1, Appl1
8	554	37.3	595	2	US-08-425-069-4 Sequence 4, Appl1
9	554	37.3	595	2	US-08-317-844B-4 Sequence 4, Appl1
10	520	35.0	1177	1	US-07-609-716-31 Sequence 31, Appl1
11	520	35.0	1177	1	US-08-175-155-29 Sequence 29, Appl1
12	520	35.0	1177	1	US-08-477-509B-64 Sequence 64, Appl1
13	520	35.0	1177	2	US-08-707-237A-35 Sequence 35, Appl1
14	520	35.0	1177	3	US-08-482-085B-64 Sequence 64, Appl1
15	520	35.0	1177	4	US-08-475-411A-31 Sequence 31, Appl1
16	520	35.0	1177	4	US-08-478-029A-31 Sequence 31, Appl1
17	507	34.1	1059	2	US-08-175-155-48 Sequence 48, Appl1
18	507	34.1	1059	2	US-08-707-237A-54 Sequence 54, Appl1
19	507	34.1	1101	3	US-08-477-509B-83 Sequence 83, Appl1
20	507	34.1	1101	3	US-08-482-085B-83 Sequence 83, Appl1
21	488	32.9	1332	1	US-07-609-716-41 Sequence 41, Appl1
22	488	32.9	1332	4	US-08-475-411A-41 Sequence 41, Appl1
23	488	32.9	1332	4	US-08-478-029A-41 Sequence 41, Appl1
24	480.5	32.4	318	4	US-09-060-756-727 Sequence 727, Appl1
25	476.5	32.1	1038	1	US-07-609-716-36 Sequence 36, Appl1
26	476.5	32.1	1038	4	US-08-475-411A-36 Sequence 36, Appl1
27	476.5	32.1	1038	4	US-08-478-029A-36 Sequence 36, Appl1

28	475	32.0	766	1	US-08-175-155-53 Sequence 53, Appl1
29	475	32.0	766	1	US-08-477-509B-88 Sequence 88, Appl1
30	475	32.0	766	2	US-08-707-237A-61 Sequence 61, Appl1
31	475	32.0	766	3	US-08-482-085B-88 Sequence 88, Appl1
32	475	32.0	979	1	US-08-477-509B-89 Sequence 89, Appl1
33	475	32.0	979	3	US-08-482-085B-89 Sequence 89, Appl1
34	475	32.0	1050	1	US-08-175-155-54 Sequence 54, Appl1
35	472.5	31.8	784	1	US-07-609-716-48 Sequence 48, Appl1
36	472.5	31.8	784	4	US-08-475-411A-48 Sequence 48, Appl1
37	472.5	31.8	784	4	US-08-478-029A-48 Sequence 48, Appl1
38	470.5	31.7	745	4	US-09-060-756-728 Sequence 728, Appl1
39	469	31.6	870	2	US-09-010-928B-28 Sequence 28, Appl1
40	469	31.6	870	2	US-09-010-928B-28 Sequence 28, Appl1
41	468.5	31.5	649	1	US-07-609-716-49 Sequence 49, Appl1
42	468.5	31.5	649	4	US-08-475-411A-49 Sequence 49, Appl1
43	468.5	31.5	649	4	US-08-478-029A-49 Sequence 49, Appl1
44	468	31.5	945	1	US-08-089-862-6 Sequence 6, Appl1
45	468	31.5	945	1	US-08-587-533-13 Sequence 13, Appl1

ALIGNMENTS

RESULT 1
US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himan, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match 68.4%; Score 1015; DB 1; Length 718;
Best Local Similarity 59.2%; Pred. No. 2.8e-75;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;
11 GSMASGNGIGGCGAGAAA-----AAAAAGGAGGGGCG 45

```

Db 93 GSGAGRGGLGGGAGAAAAAGAGGCGGCGAGAAAAAGAGCGGCG 152
OY 46 GIGSGCTGSRGGLGGGAGAAAAAGAGGCGGCGAGAAAAAGAGCGG 96
Db 153 GIGSGG -AGRGGLGGGAGAAAAAGAGGCGGCGGAGGAGGAGGAG 210
OY 97 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 135
Db 211 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 266
OY 136 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 182
Db 267 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 325
OY 183 -AGAGAGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 220
Db 336 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 384
OY 221 -AGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 259
Db 385 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 443
OY 260 -AGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 279
Db 444 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 485

RESULT 2
US-08-317-844B-2
; Sequence 2, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Himman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317, 844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-844B-2
```

```

Query Match 68.4%; Score 1015; DB 2: Length 718;
Best Local Similarity 59.2%; Pred. No. 2.8e-75;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

OY 11 GSMASRGGLGGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGG 45
Db 93 GSGAGRGGLGGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGG 152
OY 46 GIGSGCTGSRGGLGGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGG 96
Db 153 GIGSGG -AGRGGLGGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAG 210
OY 97 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 135
Db 211 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 266
OY 136 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 182
Db 267 GGGAGAAAAAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 325
OY 183 -AGAGAGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 220
Db 336 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 384
OY 221 -AGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 259
Db 385 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 443
OY 260 -AGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 279
Db 444 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 485

RESULT 3
US-09-034-177-3
; Sequence 3, Application US/09034177
; Patent No. 6127146
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,177
; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0486 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
```



```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,298
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 530
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,747
: FILING DATE: 14-MAR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1447-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 832 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: ORGANISM: N. clavipes
: TISSUE TYPE: minor amputate gland
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..309
: US-08-458-298-2

```

```

Query Match          45.4%; Score 674; DB 1; Length 832;
Best Local Similarity 46.4%; Pred. No. 1.1e-47;
Matches 168; Conservative 7; Mismatches 97; Indels 90; Gaps 13;

```

```

QY 15 SGRGGLGCGAGAAAAAAGAGAGGCGG-----TGSQTSRGLG--- 59
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 314 AGAGGAGGGRGAGAGAAAGAGAGGCGGAGAGAAAAAGAGAGGGRGA 373
QY 60 GCGAGAAAAAAGAGAGGCGGCGG-----SGRGLGCGAGAAAAA 108
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 374 GAGAGAAAGAGAAAGAGAGGCGGCGGAGAGAAAAAGAGAGGCGAGAG 433
QY 109 AAAAGAGGCGGCGGAGSOG-----TSRGLGCGAGAAAAAAGAGAG- 156
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 434 AAA--GAGAGGTGGGCGGAGAGAGAAAAATGAGAGGTGAGAGAGAGAGTGG 491
QY 157 -GYGGLGSGGT-----SGRGLG-GGAGAAAAA-----AAAAGAGGCGGTG 197
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 492 AGYGCGGCGGAGAGAAAAAGAGAGAGTGRGAGAGAAAGAGAGAGAGAGTGG 551
QY 198 GLGSGGT-----SGRGLGCGGAGAAAAAAGAGAGGCGGTG 235
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 552 GCGGCGAGAGAAAAAGAGAGAAAGTGRGAGAGAAAGAGAGAGGCGGTG 611
QY 236 ---GLG-----GTSRGLG-GGAGAAAAAAGAGAGGCGGCGTSGTSGTRP 286
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 612 AAGAGAAAAAAGAGAGGAGTGRGAGAGAAAAAGAGAGAGTGGCGGTGAGAGAAA 671
QY 287 AA 288
: | |
DB 672 AA 673

```

```

RESULT 6
US-08-864-038A-3
: Sequence 3, Application US/08864038A
: Patent No. 6001592
: GENERAL INFORMATION:
: APPLICANT: Kunio MAKASHIMA et al.

```

```

: TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
: TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
: TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSER: 812-5 Hirano
: STREET: Isshinden
: CITY: Tsu-city
: STATE: Mie-prefecture
: COUNTRY: JAPAN
: ZIP: 514-01
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Microsoft Windows 95
: SOFTWARE: Word Perfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/864,038A
: FILING DATE: May 28, 1997
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: JP 8-184459
: FILING DATE: 15-July-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: C. Bruce Hamburg
: REGISTRATION NUMBER: 22,389
: REFERENCE/DOCKET NUMBER: F-5610
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212)986-2340
: TELEFAX: (212)953-7733
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 738
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Pinctada fucata
: CELL TYPE: mantle epithelial cell
: FEATURE:
: NAME/KEY: peptide
: LOCATION: from 1 to 738
: IDENTIFICATION METHOD: E (by experiment)
: US-08-864-038A-3

```

```

Query Match          40.0%; Score 593.5; DB 3; Length 738;
Best Local Similarity 41.1%; Pred. No. 3.2e-41;
Matches 167; Conservative 10; Mismatches 82; Indels 147; Gaps 12;

```

```

QY 11 GSMASGRGLGCGAGAAAAAAGAGAG-----GGYGLGSGGTSG--- 54
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 113 GGMAGAGGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 172
QY 55 -----RGGLGCGAGAAAAAAGAGAGGCGGCGGCGTSGRGLGCG 99
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 173 DLFDLFDLGAALALGAGAGGAGAAAAAAGCGVGGAAAAAAGAGAGAA 232
QY 100 --GAGAAAAAAGAGAG-----GGYGLGSGGTSGRGLGCGG--AGAAAAA 150
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 233 LGGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
QY 151 GAGAGGCGGCGGCGTSGRGLGCGGAG-----AAAAAAGAGAGGCGG 203
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 292 AAGGCGGAGGCGFTG--GCGGRRGRRGRRRAAAAAAAGGCGGCGG---- 344
QY 204 TSGRGLGCGAGAAAAAAGAGAG----- 226
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 345 ---GGGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 227 -----GAGCGGCGGGLGSGGTSGR--- 245
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 401 STKSQIDDLKVLKDLGLKSSASASASASASACGGGGGGGNGGGGGGAGALA 460
OY 246 -----GGGGGAGAAAAAAGGAGCGCTGGGCSQGTSG 282
      |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 461 AALAAAGAGGGGCGGGGGGALAAALAA- GAGGGGPGGGLGGLG 505

```

RESULT 7

```

US-08-529-1908-1
: Sequence 1, Application US/08529190B
: Patent No. 5833991
:
: GENERAL INFORMATION:
: APPLICANT: Masucci, Maria G.
: TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES
: TITLE OF INVENTION: CONFERRING INVISIBILITY TO THE IMMUNE SYSTEM
: NUMBER OF SEQUENCES: 76
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Witcoff, Ltd.
: STREET: One Financial Center
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02111
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Wordperfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/529,190B
: FILING DATE: 15-SEP-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: SE9501324-9
: FILING DATE: 10-APR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US08/522,595
: FILING DATE: 01-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Williams, Ph.D., Kathleen A
: REGISTRATION NUMBER: 34,380
: REFERENCE/DOCKET NUMBER: 3255/53015
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-345-9100
: TELEFAX: 617-345-9111
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 235 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
US-08-529-1908-1

```

Query Match	37.8%;	Score 561;	DB 2;	Length 235;
Best Local Similarity	51.2%;	Pred. No. 5e-39;		
Matches 133;	Conservative 5;	Mismatches 92;	Indels 30;	Gaps 9

[illegible]

Db 164 -----ACAGGAGAGAGAGAGAGA-----GGGAGGAGAGCGGAGGAGAGAGAGAGAGAGAGA 214

QY 260 AAAAAAGAGCGCGTGGGAG 279

Db 215 GAGGAGGAGAGGAGGAGAGG 234

RESULT

US-08-425-069-4
Sequence 4, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himman, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTRAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-4

Query Match 37.3%; Score 554; DB 1; Length 595;
 Best Local Similarity 44.1%; Pred. No. 4.1e-38;
 Matches 166; Conservative 15; Mismatches 101; Indels 94; Gaps 17.

QY	3	GSNNHHHNGSMASGRGGLGCGACGCAAAAAAAAAAAGCAGC	----	CGYC	-----	GLSGQ	50
Db	87	GGYGRDQGGPGGYGGQGGQPSGPGSAAAAASAESGGQPGGPGPGQGGPGGYGPGGQ					146
QY	51	GTSGRG----	GLGCGGAGCAAAAAAAAAAGAGCGGCGGLGSGDTSGRG----	GLGCGGAG			102
Db	147	GGCGYGPQGGPGSSGPGSAAAAAASGPGQDGPGGYG	----	CGQGGPGYGPQGGPGSPGSA			205
QY	103	AAAAAAAAAAGAGCGGCTGCGTSGRG----	GLGCGGACAAAA	-----			144
Db	206	AAAAAAAAAGPGQGGPGGYG	----	GQDGPGGYGPQGGPGSAAAAAAGPGQDGPGGYG			264
QY	145	-----	AAAAAAGCGGGYTG	----	GLGSGDTSGRGGLGCGGACAAAAAAGGAGQ		193
Db	265	GQGGSGPGSAAAAAANGPGGYGPQGGPGGYGPQGGPGSAGSAAAAAAGPGQGG					324
QY	194	GGYGGGLGSGDTSGRG----	GLGCGGAGCAAAAAAAAA	----	GGAGCGGCGGLGSGDTSGRG	----	246

Db 325 GGYGP-GGQGPGGYGPQQGPGGYSASAAAAAAGPQQGPGGYGP-GGQGPSPGSA 382
QY 247 -----GLGGGAG-----AAAAAAAAAG-----GAGGGYGG 274
Db 383 SAAAAAAGPBGYPGQGGPGGYAPGQGGSPGSPASAAAAAAGPBGYPGQGGPGG 442
QY 275 L--GSQGTSGIRRPAA 288
Db 443 YAPGQGGSPGSPASAA 458

RESULT 9

US-08-317-844B-4
; Sequence 4, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Himan, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELETYPE: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-844B-4

Query Match 37.3%; Score 554; DB 2; Length 595;
Best Local Similarity 44.1%; Pred. No. 4,le-38;
Matches 166; Conservative 15; Mismatches 101; Indels 94; Gaps 17;

QY 3 GSHNNHNSMASRGGLGGGAGAAAAAAGAGAGC---GGYG-----GLGSO 50
Db 87 GGYGPRGQGGPGYGPQQGPGGSAASASASGQGGPGYGPQQGPGGYPGQ 146
QY 51 GYSGRG-----GLGGGAGAAAAAAGAGAGCGTGGGSGTSGRG-----GLGGGAG 102
Db 147 GPGGYGPQQGPGSPGSAASAAAAAGPQQGPGGYGP-GQGGPGGYGPQQGPGSPGSA 205
QY 103 AAAAAAAGAGAGGAGGGLGSGTSGRG-----GLGGGAGAAAA-----144
Db 206 AAAAAAGSGGQGGPGGYGP-GQGGPGGYGPQQGGLSGPSASAAAAAAGPQQGPGGYGP 264

QY 145 -----AAAAAGAGGCGYGP-GLGSGTSGRGGLGGGAGAAAAAAGAGAGC 193
Db 265 GGQGPSPGSAASAAAAAAGPBGYPGQGGPGGYGPQQGPGGASAAAAAAGPQQGL 324
QY 194 GGYGGLGSGTSGRG-----GLGGGAGAAAAAAGAGAGCGTGGGSGTSGRG-- 246
Db 325 GGYGP-GGQGPBGYPGQGGPGGYGPGSASAAAAAAGPQQGPGGYGP-GGQGPSPGSA 382
QY 247 -----GLGGGAG-----AAAAAAAAAG-----GAGGGYGG 274
Db 383 SAAAAAAGPBGYPGQGGPGGYAPGQGGSPGSPASAAAAAAGPBGYPGQGGPGG 442
QY 275 L--GSQGTSGIRRPAA 288
Db 443 YAPGQGGSPGSPASAA 458

RESULT 10

US-07-609-716-31
; Sequence 31, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-609-716-31

Query Match 35.0%; Score 520; DB 1; Length 1177;
Best Local Similarity 43.9%; Pred. No. 4,le-35;
Matches 122; Conservative 22; Mismatches 128; Indels 6; Gaps 4;

QY 11 GSWASGRGLGGGAGAAAAAAGAGAGCGYGLGSGTSGRGGLGGGAGAAAAA 70
Db 71 GSGAAGYAG 130
QY 71 AAAAG-GAGGAGYGLGSGTSGRGGLGGGAGAAAAAAGAGAGCGYGLGSGTGS 129
Db 131 SGAAAGYAG 190
QY 130 GRGGLG---GGGAGAAAAAAGAGAGCGYGLGSGTSGRGGLGGGAGAAAAA 186


```

0Y      71 AAAG-CAGCGGTGGTCSQSTSGRGGLGCGACAGAAAAAAAGAAGGAGGCGGTGGTSG 129
Db      131 :||||| |:||| |: ||| |: | |: ||| |: ||| |: ||| |: ||| |: ||| |: ||| 130
OY      130 GRCGLG---CGAGAAAAAAAAAAGCAGCGCGTGCLGSQCTSGRGLGCGACAGAAAAAAA 186
Db      191 GAACGYCAGAGSAGAGSCGACAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGS 250
OY      187 AAG-GAGCGGTGCLGSQCTSGRGLGCGAGCAAAAAAAAAAGCAGCGGTGGTCSQSTSCR 245
Db      251 AAGYGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSGA 310
OY      246 GGLG-CGAGCAAAAAAAAAAAGCAGCGGTGGTGGTCSQSTSG 282
Db      311 AGYGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 348

RESULT    13
US-08-707-237A-35
; Sequence 35, Application US/08707237A
; Patent No. 5830713
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Capello, Joseph
; APPLICANT: Crissman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
; TITLE OF INVENTION: REPETITIVE DNA
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/707/237A
FILING DATE: 03-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-10/WHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEX: (415) 398-3249
TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 35:

```

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-707-237A-35

Query Match      35.0%; Score 520; DB 2; Length 1177;
Best Local Similarity 43.9%; Pred. No. 4.1e-35;
Matches 122; Conservative 22; Mismatches 128; Indels 6; Gaps 4;

QY 11 GSMSASGRGCLGCGAGAAAAAAGAGAGGGYGGTSGTSGRGLGCGAGAAAAA 70
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 71 GSGAGYAGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 130
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 71 AAAAG-GAGGGYGGTSGTSGRGLGCGAGAAAAAAGAGAGGGYGGTSGTSG 129
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 131 SGAAAGYAGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 190
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 130 GRGGLG---GCGAGAAAAAAGAGAGGGYGGTSGTSGRGLGCGAGAAAAA 186
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 191 GAAGYAGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 250
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 187 AAG-GAGGGYGGTSGTSGRGLGCGAGAAAAAAGAGAGGGYGGTSGTSGR 245
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 251 AAGYAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 310
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 246 GGLG-GCGAGAAAAAAGAGAGGGYGGTSGTSG 282
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 311 AGYAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 348

RESULT 14
US-08-482-085B-64
Sequence 64, Application US/08482085B
Patent No. 6018030
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Richardson, Charles
APPLICANT: Chambers, James
APPLICANT: Causey, Stuart
APPLICANT: Pollock, Thomas J.
APPLICANT: Cappelletto, Joseph
APPLICANT: Crissman, John W.
TITLE OF INVENTION: NO. 6018030e1 Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,085B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049

```

```

?      FILING DATE: 22-APR-1993
?      PRIOR APPLICATION DATA:
?      APPLICATION NUMBER: US 08/175,155
?      FILING DATE: 29-DEC-1993
?      ATTORNEY/AGENT INFORMATION:
?      NAME: Trecaartin, Richard F.
?      REGISTRATION NUMBER: 31,801
?      REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTK
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: 415-781-1989
?      TELEFAX: 415-398-3249
?      INFORMATION FOR SEQ ID NO: 64:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1177 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: peptide
?      US-08-482-085B-64

```

Query Match	35.0%;	Score 520;	DB 3;	Length 1177;
Best Local Similarity	43.9%;	Pred. NO. 4.1e-35;		
Matches 122;	Conservative 22;	Mismatches 128;	Indels 6;	Gaps 4;

QY	11	GSMASGRGLGCGAGAAAAAAAAGGAGCGCGYGGTGGTSGTSGRGLGGGCGAGAAAAA	70
Db	71	GSGAAGTGGAGAGSSAGAGSCGAGAGSGAGAGSGAGAGSGACAGSGAGAGSGAGAG	1300
QY	71	AAAG-GAGCGGCGTGSQGTSGRGLGGCGCAGAAAAAAAAGGAGCGGCGTGS	139
Db	131	SGAAGTCAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	180
QY	130	GRGGLG---GGGACAAAAAAAAGGAGCGGCGTGSQGTSGRGLGGCGAGAAAAAA	186
Db	191	CAACTGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGC	230
QY	187	AAG-GAGCGGCGTGSQGTSGRGLGGCGCAGAAAAAAAAGGAGCGGCGTGSQGTSGR	245
Db	251	AAGTCAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA	310
QY	246	CGLG-CGGACAAAAAAAAGGAGCGGCGTGSQGTSG	282
Db	311	AGTCAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA	348

RESULT 15
 US-08-475-411A-31
 Sequence 31, Application US/08475411A
 Patent No. 6140072
 GENERAL INFORMATION:
 APPLICANT: Ferrari, Franco A.
 APPLICANT: Cappello, Joseph
 TITLE OF INVENTION: Functional Recombinantly Prepared
 TITLE OF INVENTION: Synthetic Protein Polymer
 NUMBER OF SEQUENCES: 119
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: CA
 COUNTRY: US
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/475,411A
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:

```

1      APPLICATION NUMBER:  US 07/609,716
2      FILING DATE:  06-NOV-1990
3
4      PRIOR APPLICATION DATA:
5
6      APPLICATION NUMBER:  US 07/269,429
7      FILING DATE:  09-NOV-1988
8
9      PRIOR APPLICATION DATA:
10
11     APPLICATION NUMBER:  US 07/114,618
12     FILING DATE:  29-OCT-1987
13
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER:  US 06/927,258
16     FILING DATE:  04-NOV-1986
17
18     ATTORNEY/AGENT INFORMATION:
19
20     NAME:  Trecaartin, Richard F.
21     REGISTRATION NUMBER:  31,801
22     REFERENCE/DOCKET NUMBER:  A-55186-9/RFT/MTK
23
24     TELECOMMUNICATION INFORMATION:
25
26     TELEPHONE:  415-761-1989
27     TELEFAX:  415-398-3249
28
29     INFORMATION FOR SEQ ID NO:  31:
30
31     SEQUENCE CHARACTERISTICS:
32
33     LENGTH:  1177 amino acids
34
35     TYPE:  amino acid
36
37     STRANDEDNESS:  single
38
39     TOPOLOGY:  linear
40
41     MOLECULE TYPE:  protein
42
43     US-08-475-411A-31

```

Query Match	35.08;	Score 520;	DB 4;	Length 1177;
Best Local Similarity	43.98;	Pred. No. 4.1e-35;		
Matches 122;	Conservative 22;	Mismatches 128;	Indels 6;	Gaps 4;

[illegible]

Search completed: July 3, 2001, 14:56:27
Job time: 517 sec

Nature 393. 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: E70806
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1381 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17744.1; PID:g9292444
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3507
i:Superfamily: collagen alpha 1(IV) chain

Query Match	36.7%	Score 544.5	DB 2	Length 1381
Best Local Similarly	37.8%	Pred. No. 1.1e-25		
Matches 144	Conservative 17	Mismatches 115	Indels 105	Gaps 13

```

Qy      1 MRGSHHHHHGSMASGRGGLGQAGAAAAAAAAGAGGQGGYGL--GSQGTSGRGL 58
      | | : : : | | | | | | | | | | | | | | | | | | | | | |
Db    205 MGSCTGCGNGCAGALLLIGGGGL--GAGCGMTGGCTGCTGCGNGCAGALLIGAGVGAGGI 262

```

QY 59 GGCGAGAAAAAAGGAGGGGCTGTGTCGGTGGGAGCAAAAAAAAAAGA --- 115
||||| || | | | | | | | | | | | : ||||
Db 263 GGGCTGAGCAAGAGCTG - GAGCAGAGTGEMNG -- GDGAGAGGGGDCGAAGDAAAASAGCTGK 318

```

QY 116 -----GGCGYGLGSG-QTSGRGLGGQC-----AG 140
      | | | | | | | | | |
Db 319 GGCGGDCGTGAGAGAPVLFHGAGAGCGGGCGTGMGAGAGDCTTAACTGGEGCTGG 378

```

QY 141 AAAAAAAAAAGCAGCGG--YGLGSQGTSGRGIGTGAGAAAAAAAAG---GAGGG 194
||| ||| | :|: ||| ||| ||| ||| :|
Db 379 AAGACGAAGRCAALTSGLAGCVGAGCTGCTGCNGNDAAAVVGFGANDDPGFACGGK 438

	QY	195	GYGHSQ-----GTSGRGSLGQ-----GAGAAAAA	221
		: :	: :	
		: :	: :	
	Dd	439	GNGGTGGAAATTCGVADGDGTGGKGGTGCAGACAGCNDAGSTCNPGKCGCDGCTGCAGAGAGA	498

QY 222 AAAA----ACGAGGGGCGGCGTSGRGLGCG-----GAGAAAAAA 261
| | | | | : | | |
Db 499 AGTCGCGHACGTGGCGDGGTGCNGNGTGGVNCADNTLPDTPCGAGCEPCCGAGGAGG 558
| | | | |

QY	262	AAAGAGGCGTGGTSG	282
		:	
DB	559	AAGCGGTGGTGGNGGNG	579

RESULT 8
B70807

C:Accession: B70807
C:Species: Mycobacterium tuberculosis
C:hypothetical glycine-rich protein Rv312 - Mycobacterium tuberculosis (str
C:date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

R. Cole, S. T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, J.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Ratajczak, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, N. *Nature* 393, 537-544, 1998.

A:Authors: Sgares, R.; Suston, J.E.; Taylor, K.; Whitehead, S.; Barrell,
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the co
A:Reference number: A70500; MUID:98295987
A:Accession: 570807

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1079 <COL>

A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3512

Query Match 36.6%; Score 543.5; DB 2; Length 1079;

Best Local Similarity 40.78; Pred. No. 1.1e-25;
Matches 144; Conservative 12; Mismatches 109; Indels 89; Gaps 14;

```

14 ASGRGGLGGGGA-GAAAAAAAAAAGAGAGGGGGLGSGQGT-----GRG 56
   | : | | | | | | | | | | | | | | | | | | | | | | | | | |
217 ANGAGNGDGGDGAAGAVGTSGATGACDGGHGCTGAAAGAGCTGAGAGSGIDVGSGT 276

```

y
57 GUGGQAGAAAAAAGAG-QGGYGLGSQ-----GTSGRGLGGQG- 100
| | | | | : | | | |
277 GTGGNGCNAIGAGCDAGSGNSGNGITGKKGNNAGAGCAAGSNGTGVANGTGCGDG 336

101 -AGAAAAAAGAGAGGCGTGGGCGTSGRGLGGG-----AGAAAAA 148
 ||| ||| : |||| | | | |||| | | |
 337 NGCAGAAATAGSNGCAGTSAAGNG--GTGGRGSGGAGAGGIGGVGCGKGNADGEVG 394

149 AAGCAG-----GGCYGGLSQGTSGKRGGLGGCGAGAAAAAAGAGAGCGCYGCL-- 199
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
395 GAGCAGGSGPNTSPGNGGOGGOGSGCAG-GAAAGACAGCGANGTGMNGGOGGATGTTGG 453

200 -----GSGGTSGRCGLGGCGAGAAAAAAGAGAGCGGTGCLGSQGTSGRG 246
 | : | | | | | | | | | | | | | | | | : |
454 AGAASSATNGSGGAGGTGCGGSSGAGCTGCAGTCGAACDGGCGCGAGC-GAGCGC 512

247 GGGGG-----AGAAAAAAAAAGCA-GGGGY-GGLGSGTSG 282
 ||| | ||| ||| ||| ||| ||| :
 513 GAGCAGTCGNGNGNITGTGTAGTAGAAGNGGACGAGGGCGTGGCGCGGGAG 566

RESULT 9

Accession: H70845
Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
Species: Mycobacterium tuberculosis
Hypothetical glycine-rich protein Rv3345c - Mycobacterium tuberculosis (strain H37Rv)

Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; andream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares
ture 393 537-544 1998

Authors: Squires, R.; Suston, J.E.; Taylor, K.; Whithead, S.; Barrell,
Title: Deciphering the biology of *Mycobacterium tuberculosis* from the co
Reference number: A70500; MUID: 98295987
Accession: 070845

;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-1538 <COL>

Experimental source: strain H37Rv
Genetics:
Gene: RV3345c

Query Match	Score	DB 2	Length
36.28	538	1538	

11 GSMASGRGGLGGGCGACAAAAAAGAGAGGGYGLSGGCTSGRGGLGGGACAAAAA 70
Matches 126; Conservative 13; Mismatches 131; Indels 54; Gaps

1030 GMINGGLGFCGACGGGAVDVAAATTGAGAGNGAGCGFASTGLGCGPGAGGPGGAGDFASG 108
71 AAAAGG-----GCGGTCGLSGG-----TSGRGLGGGCGCAAAAAAAAAAGC----- 114

1090 VGVVGGAGCGDGGAGCGVGVGGGGGIGGEGRTGNGSGSGDGGGIGLGGNGLGGNGVVS 114
115 -----ACGGGVGGLGSGGTGGRGGLGGGAGAAAAAAGAGACGGCGVGLGTSQ--- 164

1150 ETGFGAGCGNGYGGPC--GPEGNGGLGCGNGAGCGNGSVTTGGDGAAGCKGNGGDGNN 1200

165 -----GTSRGRGLGCGGAGAAAAAAGG-----AGCGGYGLSQGTS 205

1208 VGLGCDAGSSGAGGNGCIGTDCAGAGAGAGGNGSSKSTTTGNAGSSGAGGNGTGCLN 1266

Db 1268 GAGGAGGAGGAGNAGYGVFEGNAAGCGDGGGNGGNGGCHGGDGTGGAGAGKGNGSSGAASGG 1327

QY 262 ---AAAGGAGCGGTGGTGGAGTSG 282

1328 VVNYTACGGGCGGCGGCGGCGGCGSAG 1351

Db

RESULT	10
A70812	
hypothetical	glycine-rich protein Rv0833 - Mycobacterium tuberculosis (strain H37RV)

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70812
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajadurai, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-749 <COL>
A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAI7639.1; PID:g2916688
A:Experimental source: strain H37Kv
C:Genetics:
A:Gene: RV0833
A:Superfamily: elastin

Query Match	36.08;	Score 535;	DB 2;	Length 749;
Best Local Similarity	39.98;	Pred. No. 2.6e-25;		
Matches 133; Conservative	9;	Mismatches 123;	Indels 68;	Gaps 11;

```

0y      16 GGGGCGGCACAAAAAAGCAGCGGTGGGSG-----GTSCR 55
          | | | | | | | | | | | | | | | | | |
Db       3 GNGCGGSGCAPCAICGACGPAGLIGVGAGGAGGDSAVAGVIGAGAGGAALLFCAGA 62

```

```

07 56 GGLGCGGAGAAAAAAGCA-----GQGGYGGLSGQGTSGRGGAGGGGAGAAAAA 10
    || || | : || | || | || | || | || | :
db 63 GGAGGSG-GGGAAGCAGCAGAGGAGGLFASGGGSGGFGGFASTGTGAGAGTGGAGGLFASGCV 12

```

```

09      AAAAGCAGCGGCGGLGC-----TSGRGGCGGAGAAAAAAGCAGCGGGL 16
       ||||| ||:::| | : ||||| | | | | | | | | | | | | | | | |
Db 122 GGTGCGAGSAGCTGGVGGTCGAGGLFASGCAAGCAGSAGCTGGAAGTCGAGGLFAGAGAGGL 18

```

Db 182 GGCGNHTGCHGAGCSAGLLALBDGAGSAGAAATTGTGACGAGCAKGLLFFSGAGGS 24

Db 242 GGAAGTGGDTCNSGAGGAGKAGLLFGSGAGGSGCAGFANGSTGAGGAGGAGLLG 30

Db 302 NGGCGSGGTSVATGGAGNGGAGAG--GGAGL 332

hypothetical, glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37RV)

C:Date: 17-Jul-1998 #sequence:revision 17-Jul-1998 #text:change 20-Jun-2000
C:Accession: A70869
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987
A:Accession: A70869
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1660 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:g2361507; PIDN:CAA16067.1; PID:g2799
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2490c
C:Superfamily: collagen alpha 1(IV) chain

Query Match	35.4%;	Score 525.5;	DB 2;	Length 1660;
Best Local Similarity	40.7%;	Pred. No. 1.7e-24;		
Matches 137; Conservative	12;	Mismatches 117;	Indels 71;	Gaps 13;

```

Qy 14 ASGAGGGLGG-GAGAAAAAAGCAGGGCGGLGSGGTSGR-----GSLGGGCA 63
      | | | | | : | | | | | : | | | |
Db 1022 AGCAGCAGCAGCAGGGSVSGDGGAGGNGAGGNGVGVCASGAGARGANGIDISGTTGACG 1081

```

```

          1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
QY      64 GAAAAAAAAAGAGAGGGGCG-----LGSQSTSGRGLGGGAGAAAAAAAAAGG--- 114
          | | | | | | | | | | | | | | | | | | | |
Db     1082 GGGGCGAGCGVGGHGGDGGCGVGAAPSGTGVGSHGTGGVGGDGLG-GAGCGVGGAGCGNGGIGI 1140

```

```

07      115  -ACGGGCGTGSQ--GISGKGGTGGQ-----GAGAAAAAA 14
          | | | | | | | | | | | | | | | | | | | | | |
Db     1141  TVGAGAGAGNGGDPCAGGKRGLTGDSGNCTSAANGYDASKHCPPLTGDDGVGNGAKAA 1200

```

[illegible]

Db 1259 GDSGKGDDGAHGAALGGLTVAGCGNGGAGCAGGAGGAGGAGFLGGCGNGGAGGGGAGGAGGSSP 131B

Db 1319 GGGGGVGGHGGAGAGDAGMNGSGGTGGGGGNGAAGGAG 1355

RESULT 12
A26099
glycine-rich cell wall structural protein - garden petunia
(*G. grandiflora*, Petunia x hybrida (various petunia))

CiDate: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A26099
R:Condit, C.M.; Meagher, R.B.
Nature 323, 178-181 1985

A: Molecule type: mRNA
A: Accession: A26099
A: Reference number: A26099
Affiliate: A gene encoding a novel glycine-rich structural protein of pecunia

C: Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Best Local Similarity 43.6%; Pred. No. 6,3e-25;
Matches 127; Conservative 4; Mismatches 126; Indels 34; Gaps

Db 60 SFGKRGAGCGGFCGCGAGGAGGCGGCGGCGGAG-----GGGCGG-CGGAGGGFCG 113

Db 114 GAGCGAGGCGGGGCGGCGGAGGGGCGAGGGGCGAGGGGAGAGAGGGGGG 173

Db 174 GFGCGGGGGGSGCHGCGFCAAGGCGGAGGGLGGCGGGGGGGGGGTC--GSSGH 231

This Page Blank (uspto)


```

FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 535 21.
FT REPEAT 536 556 22.
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT REPEAT 662 662
FT CONFLICT 672 672
FT CONFLICT 672 672
FT CONFLICT 695 747

```

SEQUENCE 747 AA: 60528 MW: 850E44B0D649E012 CRC64:

Query Match 68.4%; Score 1015; DB 1; Length 747;
 Best Local Similarity 59.2%; Pred. No. 3.8e-43;
 Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

```

QY 11 GSASGSGGGLGGGAGAAAAA-----AAAAAGAGGGGGY 45
DB 93 GSGGAGRGGLGGGAGAAAAAAGAGGGGGLGNOGAGRGGQGAAGAGGGGGY 152
QY 46 GLGSGGTSGRGGLGGGAGAAAAA-----AGAGGGGGLGSGGTSGRGGL 96
DB 153 GLGSGG-AGRGGLGGGAGAAAAAAGAGGGGGLGNOGAGRGGGLGSGG-AGRGGL 210
QY 97 GGGGAGAAAAA-----AAAAAGAGGGGGLGSGGTSGRGGL 135
DB 211 GGGGAGAAAAAAGAGGGGGLGSGGAGAGASAAAAAGAGGGGGLGSGG-AGRG 266
QY 136 GGGGAGAAAAA-----AGAGGGGGLGSGGTSGRGGLGSGGAGAAAA- 182
DB 267 GGGGAGAAAAAAGAGGGGGLGSGGAGGGGGLGSGG-AGRGGLGGGAGAAAAAGAG 325
QY 183 -----AAAAAGAGGGGGLGSGGTSGRGGLGGGAGAAAA----- 220
DB 326 GGGGAGAGGAGAAAAAAGAGGGGGLGSGG-AGRGGLGGGAGAGVAAAAAGAGG 384
QY 221 -----AAAAAGAGGGGGLGSGGTSGRGGLGGGAGAAAA- 259
DB 385 GGGGGLGSGGAGRGGAGAGAAAAAGAGGAGGAGGGLGNOG-AGRGGLGGGAGAAAA 443
QY 260 -----AAAAAGAGGGGGLGSGG 279
DB 444 GAGAGGGGGLGNOGAGRGGAGAAAAAGAGGGGGLGSGG 485

```

RESULT 2
 EBN1_EBV STANDARD: PRT: 641 AA.
 AC P03211;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE EBNA-1 NUCLEAR PROTEIN.
 GN BKRF1.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;

BAER R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 Gibson T.J., Hatfull G., Hudson G.S., Sachwell S.C., Seglin C.,
 Tufnell P.S., Warrall B.G.;
 "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 Nature 310:207-211(1984).
 [2]
 RP SEQUENCE OF 1-26 FROM N.A.
 RX MEDLINE=86259739; PubMed=3460083;
 RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
 "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
 proteins: a probable transcriptional initiation site.";
 Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
 [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=90266473; PubMed=2161150;
 RA Petri L., Sample C., Kieff E.;
 "Subnuclear localization and phosphorylation of Epstein-Barr virus
 latent infection nuclear proteins.";
 Virology 176:563-574(1990).
 [4]
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
 RP MEDLINE=96006523; PubMed=7553871;
 RA Bochkarev A., Batwell J.A., Pinetzer R.A., Furey W.F., Jr.,
 Edwards A.M., Freppier L.;
 "Crystal structure of the DNA-binding domain of the Epstein-Barr
 virus origin-binding protein EBNA 1.";
 Cell 83:39-46(1995).
 CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
 MAINTENANCE REPLICATION OF EBV EPISOME. "TRANSACTIVATING FACTOR
 FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORP."
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT
 ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
 WITH THE NUCLEAR MATRIX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL: V01555; CAA24816.1; -.
 DR EMBL: M13941; AAA45889.1; -.
 DR PIR: A03773; Q0BE31.
 DR PIR: S33021; S33021.
 DR PDB: 1VHI; 23-DEC-96.
 DR TRANSFAC: T00211; -.
 KW Nuclear protein; DNA-binding; Transcription regulation; Activator;
 KW 3D-structure.
 FT DOMAIN 87 352 GLY/ALA-RICH.
 SO SEQUENCE 641 AA: 56427 MW: 4D161653E16FC341 CRC64;

Query Match 41.1%; Score 610.5; DB 1; Length 641;
 Best Local Similarity 48.6%; Pred. No. 9.1e-24;
 Matches 143; Conservative 14; Mismatches 104; Indels 33; Gaps 10;

```

QY 2 RGSNNHHHNSMAGSGGGLG-GGGAGAAAAAAGAGAGG-GGYGGLGSGGTSGRGGLG 59
DB 83 KGTN-----GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGA 130
QY 60 GGGAGAAAAAAGAGAGAGG-GGYGGLGSGGTSGRGGLGGGAGAAAAAAG--GA 115
DB 131 GAGGAGAGAGAGAGAGAGAGAGAGAGAG--GAGGAGAGAGAGAGAGAGAGG 188
QY 116 GGGGGLGSGGTSGRGGLGGO-GAGAAAAAAGAGAGGGGGLGSGGTSGRGGLG 174
DB 189 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG--AGGAGAGG 244
QY 175 GAGAGAAAAAAGAGAGGGGGLGSGGTSG-----RGGLGGGAGAAAAAAGAGG 230
DB 245 AGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 304

```


[illegible][illegible]

	RESULT	5
ID	GRPL_PETHY	
AC	P09789	
NCBI	TaxID=4102;	
OC	Solanales; Solanaceae; Petunia.	
CC	-1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).	
CC	-1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).	
CC	-1- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.	
CC	-1- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.	
CC	-1- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40 AA.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - -	
CC	between the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isdb-sib.ch/announce/ or send an email to license@isdb-sib.ch).	
CC	-----	
DR	EMBL; X04335; CAA27866.1; -.	
DR	PIR; A26099; A26099.	
DR	HSSP; P30129; ADPV.	
KW	Cell wall; Structural protein; Repeat; Signal.	
FT	SIGNAL	1 27
FT	CHAIN	28 384
FT		GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1.

	RESULT	6		
YO34_MYCTU	STANDARD;	PRT;	778 AA.	
ID	YO34_MYCTU			
AC	P71933:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.			
GN	RV2634C OR MTCY441..04C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OX	NCBL_TaxId=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gae S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,			
RT	Taylor K., Whitehead S., Barrell B.G.;			
RL	"Deciphering the biology of Mycobacterium tuberculosis from the			
CC	complete genome sequence."			
CC	Nature 393:537-544(1998).			
CC	-I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on the			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z80225; CAB02341.1; -.			
DR	HSSP; P04002; IATF.			
DR	Tuberculist; RV2634C; -.			
DR	InterPro; IPRO00084; -.			
DR	InterPro; IPRO01899; -.			

[illegible]

QY	63	AGCAAAAA-----AAAGAGAGCG--YGLGSGQSTGRGGLGQAGAAAAA-----	109
DB	191	AGCGAGGGYGGGGGHEGGGGGGGGGAGAGGTGAGGEGHGAAGGGGAGGGYAGAGG	250
QY	110	AAAGAGAGGGYGGGLGSGQSTGRGGLGQAGAAAAA-----AAAGAGGGYGGGLG	162
DB	251	GGAGGGCGGAGAGGAGGAGGAGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	310
QY	163	SGQSTGRGGLGQAGAAAAA-----AGCAGGGYGGGLGSGQSTGRGGLGQAG	216
DB	311	AGGEGGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	370
QY	217	AAAAA-----AAAGAGAGGGYGGGLGSGQSTGRGGLGQAGAAAAAAGAGAGGGY	272
DB	371	AGGCGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	430
QY	273	GGLGSGQSTG 282	
DB	431	GG-AGAGGGG 439	
RESULT	9		
YD25_MYCTU	ID	YD25_MYCTU	STANDARD; PRT; 603 AA.
AC	Q10637;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	HYPOTHEICAL PE-PGRS FAMILY PROTEIN RV1325C PRECURSOR.		
DN	RV1325C OR MWCY130.10C.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Filicutes; Actinobacterii; Actinobacteridae;		
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37RV;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,		
RA	Taylor K., Whitehead S., Barrell B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RL	Nature 393:537-544(1998).		
CC	-1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS		
CC	SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; 273902; CAA98089.1; -.		
DR	HSSP; P19972; IKVD.		
DR	TubercuList; RV1325C; -.		
DR	InterPro; IPR000084; -.		
DR	Pfam; PF00934; PE; 1.		
KM	Hypothetical protein; Repeat; Signal.		
FT	SIGNAL	1	30
FT	CHAIN	31	603
FT			POTENTIAL.
FT			HYPOTHEICAL PE-PGRS FAMILY PROTEIN
FT			RV1325C.
FT			GLY-RICH.
SO	SEQUENCE	603 AA; 49575 MW; 4E9BC82807AE964 CRC64;	

Query Match 33.4%; Score 496; DB 1; Length 603;
 Best Local Similarity 37.6%; Pred. No. 2,7e-18;
 Matches 137; Conservative 9; Mismatches 118; Indels 100; Gaps 16;

16 GRGGLGCGGAGAAAAAAGCA--GCGYGLGSGTSGRGLGCGGAGAAAAA 73
 167 GAGGAGGDDGCGNGAGAGAGAGGGLFSGSGAGGPGVGNVTGGLDGG-CAAGLFGAGG 225
 74 AGGAG-----GCGYGLGSG-----GTSRGLGCGGAGAAAAA 107
 226 IGGAGGFGFNGAGAGAGAGAGGSLFEVLAAGAGGTGGLSVNGCTGTGTGG-GGGLFSG 284
 108 AAAAAAGAG-----GCGYGLGSG-----GTSRGLGCGGAGAAAAA 153
 285 GAGGAGGFGVSGSAGNGCTGCGDGIPTGNGCTGTGTGNDLVGEGGAGAGAGNAG 344
 154 ---GCGYGLGSG-----GTSRGLGCGGAGAAAAAAGAG----- 192
 345 ILFGAGGTGTGTGAGAPDPGCTGCGKGVGGIG-GAGALFPGGAGGTGCGFSSADQM 403
 193 GCGYGLGSGTSGTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
 404 AGGAGGSGGSGAGAKLIGDGGAGGTGDSVRGAAGSGGTGTGGLIDGGAGAGGTGIE 463
 242 ---TSGRGLGCGGAGAAAAAAGAGAG-----GCGYGLGSG 278
 464 FGSGVGGAGGAGNAAGLSGAGAGAGCGFETAGDGGAGNAGLINDGGAGAGGLGIA 523
 279 GTSG 282
 524 GDCG 527

RESULT 10
 WA22_MYCTU STANDARD; PRT: 914 AA.
 ID 006794;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MAG22 ANTIGEN PRECURSOR.
 GN MAG22 OR RV1759C OR MYCY28.25C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_Taxid=1773;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; Pubmed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tenor R.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagers K., Krogan A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:337-544(1998).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGSS
 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL: Z95890; CAB09322.1; -
 DR HSSP: P41140; 2SFA
 DR Tuberculin; RV1759C; -
 DR InterPro: IPR000084; -
 DR Pfam: PF00934; PE; 1.
 KW Antigen; Repeat; Signal.
 FT STGNL 1 30
 FT CHAIN 31 914
 SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBEB6AC8 CRC64;

Query Match 33.3%; Score 494; DB 1; Length 914;
 Best Local Similarity 39.4%; Pred. No. 4.3e-18;
 Matches 134; Conservative 10; Mismatches 126; Indels 70; Gaps 15;

16 GRGGLGCGGAGAAAAAAGCA-----GCGYGLGSGTSGRGLGCGGAG 65
 164 GNGAGAGAGGTATANNNGAGAGAGAGSMLFAGAGAGAGAGAAATSLVGIGTGTGNA 223
 66 AAAAAAAGAG-----GCGYGLGSG-----TSGRGLGCGG-AGAAAAA 112
 224 GMLAGAGAGAGAGGFSSTAGAGAGAGAGGLFTTGAGAGAGAGAGAGAGAGAG 283
 113 GGAGGGYGGYGLGSGTSGRGLGCGGAGAGAA--AAAAAAGAG----- 154
 284 GAGMGAGAGGFGDHTGTGTGAGAGDGGGGLFGAGGAGAGAGAGAGAGAGAGAG 343
 155 ---GCGYGLGSGTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201
 344 SLGAGAGAGGTGAGAGGTGFGGKAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
 202 ---GCTSGRGLGCGG--AGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 252
 404 SAGMLSGSGSGGAGGSGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
 253 AGA-----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 286
 464 GNAVLINGEGEGIGLAKSGFGFGGL-LLGADGTNAP 502

RESULT 11
 GRP_ARATH STANDARD; PRT: 338 AA.
 ID P27483;
 AC 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN PRECURSOR.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 NCBI_Taxid=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLOMBIA;
 RX MEDLINE=92003708; Pubmed=1912511;
 RA Quigley F., Villiot M.L., Maché R.;
 RT "Nucleotide sequence and expression of a novel glycine-rich protein
 gene from Arabidopsis thaliana.";
 RL Plant Mol. Biol. 17:949-952(1991).
 CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: 278020; CAB01461.1; -.
 DR HSSP: P19972; IKVD.
 DR TUBerculist; RV1818c; -.
 DR InterPro: IPR000084; -.
 DR Pfam: PF00934; PE: 1.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 39 59 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 135 155 POTENTIAL.
 FT TRANSMEM 158 178 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 244 264 POTENTIAL.
 FT TRANSMEM 281 301 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 338 358 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 403 423 POTENTIAL.
 SQ SEQUENCE 498 AA; 40755 MW; 4F6F8F2482586BA CRC64;

Query Match 32.9%; Score 489; DB 1; Length 498;
 Best Local Similarity 40.1%; Pred. No. 5.3e-18;
 Matches 132; Conservative 14; Mismatches 123; Indels 60; Gaps 14;

QY 11 GSNASGRGGLGCGA-----GAAAAAAGAGAGGCGGCGLSGCTG-----GRGGLG 59
 DB 173 GNVASGAGAGGAGAGGAGGLYGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 232
 QY 60 GCGAGAAAAAAGG-----AGCGGCGGCGLSGCTG-----GRGGLG 98
 DB 233 GVG-GLAADAGDAGAGDGLFVGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 291
 QY 99 GCGAGAAAAAAGGAGGCGGCGGCGLSGCTG-----GCGAGAAAAAAGGAGAGC 155
 DB 292 VGDDGVAFELCTAGCGPGAGAGGLFVGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 351
 QY 156 GCGGCGLSGCTGSRGGLGCG-----GAGAAAAAAGAGAGGCGGCGLSGCTG-----G 206
 DB 352 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 408
 QY 207 RGLGCGGAGAGAAAAA--AAAGAGAGGCGGCGLSGCTGSRGGLGCGGAGG----- 255
 DB 409 AGGAGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 468
 QY 256 AAAAAAAGAGAGGCGGCGGCGLSGCTG 283
 DB 469 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 497

RESULT 14
 ELS_MOUSE
 ID ELS_MOUSE STANDARD; PRT; 860 AA.
 AC P54320;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ELASTIN PRECURSOR (TROPOLASTIN).
 GN ELN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC SRRATN-BALB/C; TISSUE=Lung;
 RX MEDLINE=95130069; Pubmed=7829060;
 RA Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
 RT Use of an intron polymorphism to localize the tropoelastin gene to
 RT mouse chromosome 5 in a region of linkage conservation with human
 RT chromosome 7.
 RL Genomics 23:125-131(1994).

CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
 CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
 CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
 CC INTO AN EXTENSIBLE 3D NETWORK.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
 CC -1- PTM: THE CROSSLINKS ARE MADE OF DEMINATED LYS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U08210; AAA80155.1; -.
 DR HSSP: P04002; 1WEA.
 DR MCD; MGT.95317; ELN.
 KW Structural protein; Repeat; Signal; Connective tissue.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 860 ELASTIN.
 SQ SEQUENCE 860 AA; 71955 MW; 0C0BE5AAE1ED7F1 CRC64;

Query Match 31.8%; Score 472; DB 1; Length 860;
 Best Local Similarity 38.5%; Pred. No. 4.8e-17;
 Matches 152; Conservative 15; Mismatches 100; Indels 128; Gaps 15;

QY 16 GRGGLGCGAGAAAAAAGG-----GCGGCGG 46
 DB 347 GIGGAGAGGPPAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 406
 QY 47 LGSCTGSGRGLGCGGAGAAAAAAGGAGGCGGCGLSGCTG-----SGRGLGCGGAGGAG 103
 DB 407 VCGGCGIGPGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 465
 QY 104 -----AAAAAAGAGGCGGCGGCGLSGCTG-----GTSRGLGCGGAGGAG 141
 DB 466 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 525
 QY 142 AAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 172
 DB 526 AAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 585
 QY 173 GCGAGAGAAAAA-----AAAG-----GAGGCGGCGGAGGAGGAGGAGGAGGAGGAG 203
 DB 586 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 645
 QY 204 TSGRGLGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 252
 DB 646 FGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 705
 QY 253 AG-----AAAAAAGAGGAGGCGGCGGCGLSGCTG 283
 DB 706 AGAAPPAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740

RESULT 15
 Y278_MYCTU
 ID Y278_MYCTU STANDARD; PRT; 957 AA.
 AC P56877;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV0278C PRECURSOR.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:55:35 ; Search time 95.41 Seconds
(without alignments)
403.530 Million cell updates/sec

Title: US-09-490-291-2

Perfect score: 1485
Sequence: 1 MRGSHHHHHMSMSGRGL.....YGLGSGCTSGIRPAKLN 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.16:*
1: SP-Archea:*
2: SP-Bacteria:*
3: SP-Fungi:*
4: SP-Human:*
5: SP-Invertebrate:*
6: SP-Mammal:*
7: SP-Mhc:*
8: SP-Organella:*
9: SP-Phage:*
10: SP-Plant:*
11: SP-Rodent:*
12: SP-Unclassified:*
13: SP-Vertebrate:*
14: SP-Virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1020	68.7	617	5	046172	nephila cia
2	1008	67.9	544	5	046171	nephila cia
3	675	45.5	988	5	017434	nephila cia
4	640.5	43.1	410	5	016988	araneus dia
5	617.5	41.6	2639	5	076786	araneus dia
6	593.5	40.0	738	5	002402	pinetada fu
7	556	37.4	1489	2	053559	mycobacteri
8	544.5	36.7	1381	2	053552	mycobacteri
9	543.5	36.6	1079	2	053557	mycobacteri
10	541.5	36.5	294	5	016986	araneus dia
11	538	36.2	1538	2	053395	mycobacteri
12	535	36.0	749	2	053844	mycobacteri
13	526.5	35.5	421	5	093119	antherea p
14	525.5	35.4	1660	2	053215	mycobacteri
15	524.5	35.3	390	10	09M3Y2	tritlicum ae
16	520.5	35.1	619	14	091PC9	cynomolgus
17	519.5	35.0	853	2	053439	mycobacteri
18	514	34.6	636	5	016987	araneus dia
19	514	34.6	1306	2	053775	mycobacteri

20	512.5	34.5	767	2	053435	mycobacteri
21	512	34.5	484	2	053394	mycobacteri
22	509.5	34.3	1468	5	0960B5	galleria me
23	507.5	34.2	923	2	053890	mycobacteri
24	507	34.1	610	5	09V5V8	dirosophila
25	503	33.9	491	2	006818	mycobacteri
26	503	33.9	731	2	050415	mycobacteri
27	499	33.6	714	2	053556	mycobacteri
28	498.5	33.6	882	2	053845	mycobacteri
29	495	33.3	576	2	P71664	mycobacteri
30	495	33.3	1329	2	006810	mycobacteri
31	494	33.3	667	2	053415	mycobacteri
32	492	33.1	443	5	09GUB4	galleria me
33	492	33.1	458	5	09N3U1	caenorhabd
34	490	33.0	349	10	09LSP2	arabidopsis
35	489.5	33.0	291	10	039337	brassica na
36	489.5	33.0	396	10	065450	arabidopsis
37	489.5	33.0	618	2	006801	mycobacteri
38	487.5	32.8	694	2	053212	mycobacteri
39	487	32.8	496	2	053631	mycobacteri
40	486.5	32.8	741	2	006808	mycobacteri
41	486	32.7	1011	2	P94986	mycobacteri
42	484	32.6	615	2	005806	mycobacteri
43	481	32.4	591	2	007224	mycobacteri
44	480.5	32.4	532	2	006215	mycobacteri
45	480.5	32.4	783	2	053809	mycobacteri

ALIGNMENTS

RESULT 1
ID 046172 PRELIMINARY; PRT; 617 AA.
AC 046172;
DT 01-JUN-1998 (TEMBLrel. 06, last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, last annotation update)
DE DRAGLINE SILK PROTEIN SPIDROIN 1 (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Arachnida: Metazoa: Arthropoda: Chelicerata: Arachnida: Araneae: Eukaryota: Metazoa: Arthropoda: Chelicerata: Arachnida: Araneae: Nephila.
OC Araneomorphae: Entelegynae: Araneidae: Tetragnathidae: Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae).";
RT J. Biol. Chem. 269:6661-6663(1994).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RA Beckwith R., Arcidiacono S., Stote R.;
RL Insect Biochem. Mol. Biol. 0:0-0(1998).
DR EMBL; U37520; AAC04504.1; -;
FT NON_TER 1 1
FT NON_TER 617 617
SQ SEQUENCE 617 AA; 49410 MW; 1EF0CE9269A832E2 CRC64;

Query Match 68.7%; Score 1020; DB 5; Length 617;
Best local Similarity 55.3%; Pred. No. 3,4e-63;
Matches 242; Conservative 6; Mismatches 12; Indels 178; Gaps 13;

QY 11 GSASGSGGLGCGAGCAAA-----AAAAAGAGCGGCGGLGS 49
DB 47 GSGGAGGGLGCGAGCAAAAGAGGCGGLGCGAGCAAAAGAGGCGGLGS 106
QY 50 QGTSRGGLGCGAGCAAAAGAGGCGGCGGLGSGGTSRGGLGCGAGCAAAAGAGGCGGLGS 109
DB 107 QG-AGRGSGGCGAG-----AAAAAGAGCGGCGGLGSGG-AGRGGLGCGAGCAAAAGAGGCGGLGS 160

```

OY 110 MAAG-----GAGGCGYGLGSGGTSGRGLGCGGAGAAAA-----144
    |||
DB 161 GGAAGGCGYGLGCGGAGCGGCGYGLGSGG-AGRGGLGCGGAGAAAAAAGAGCGGGLGSGG 219
OY 145 -----AAAAAAGAGCGGCGYGLGSGGT-----166
    |||
DB 220 AGCGAGAAAAAAGAGCGGCGYGLGSGGAGRGCGGAGAAAAAAGAGCGGCGGAGCGG 279
OY 167 -----SGRGGLGCGGAGAAAA-----182
    |||
DB 280 GGLGCGGAGRGGLGCGGAGAAAAAAGAGCGGGLGCGGAGAAAAAAGAGCGGCGG 339
OY 183 -----AAAAAAGAGCGGCGYGLGSGGTSGRGLGCGGAGAAAAA-----225
    |||
DB 340 LCNCGAGRGCGGAGAAAAAAGAGCGGCGYGLGSGG-AGRGGLGCGGAGAAAAAAGAGCGG 398
OY 226 -----AGAGCGGCGYGLGSGGTSGRGLGCGGAGAAAA-----261
    |||
DB 399 GGLGCGGAGCGGCGYGLGSGG-SGRGGLGCGGAGAAAAAAGAGCGGGLGCGGAGAAAA 457
OY 262 AAAGAGAGCGGCGYGLGSGG 279
    |||
DB 458 AAGAGYRGCGYGLGSGG 475

```

```

RESULT 2
O46171 PRELIMINARY: PRT: 544 AA.
ID 046171
AC 046171
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SPIROIN 1 (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN 12
RP SEQUENCE OF 449-544 FROM N.A.
RX MEDLINE=98148687; PubMed=9487707;
RA Arcidiano S., Mello C., Kaplan D., Cheley S., Bayley H.;
RT "Purification and characterization of recombinant spider silk
    expressed in Escherichia coli.";
RL Appl. Microbiol. Biotechnol. 49:31-38(1998).
RN 13
RP SEQUENCE FROM N.A.
RA Beckwith R., Arcidiano S., Stote R.;
RL Insect Biochem. Mol. Biol. 0:0-0(1998).
RN 14
RP SEQUENCE FROM N.A.
RA Beckwith R.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL: U20329; AAC38957.1; -.
FT NON_TER 1
SQ SEQUENCE 544 AA: 44107 MW; CC61B3551945615 CRC64;

```

Query Match 67.9%; Score 1008; DB 5; Length 544;
 Best Local Similarity 69.9%; Pred. No. 2e-62;
 Matches 230; Conservative 8; Mismatches 17; Indels 74; Gaps 13;

```

OY 11 GSMASGRGLGCGGAGAAAA-----AAAAAAGAGCGGCGYGLGSGG 51
    |||
DB 64 GSGAGRGGLGCGGAGAAAAAAGAGCGGGLGCGGAGGAGAAAAAAGAGCGGCGYGLGSGG 123
OY 52 TSGRGGLGCGGAGAAAAAAGAGCGGCGYGLGSGGTSGRGLGCGGAGAAAAA 111
    |||

```

```

DB 124 -AGRGSGCGGAG----AAAAAAGAGCGGCGYGLGSGG-AGRGGLGCGGAGAAAAAAGC 177
OY 112 AG-----GAGCGGCGYGLGSGGTSGRGLGCGGAGAAAA-----AA 146
    |||
DB 178 AGCGGCGYGLGCGGAGCGGCGYGLGSGG-AGRGGLGCGGAGAAAAAAGAGCGGCGGAGAA 236
OY 147 AAAAGAGCGGCGYGLGSGGTSGRGLGCGGAGAAAAA-----AGAGCGGCGYGLG 200
    |||
DB 237 AAAAGAGCGGCGYGLGSGG-----AGRGCGGAGAAAAAAGAGCGGCGGAGCGGCGYGLG 292
OY 201 SGTSGRGLGCGGAGAAAAAAGAGCGGCGYGLG-----SGTSGRGLG-----G 250
    |||
DB 293 SGG-AGRGGLGCGGAG--AAAAAAGAGCGGCGGCGGAGAAAAAAGAGCGGCGGAG 349
OY 251 OGAGAAAAAAGAGAGCGGCGYGLGSGG 279
    |||
DB 350 CGAGAAAAAAGAGYRGCGYGLGSGG 378

```

```

RESULT 3
O17434 PRELIMINARY: PRT: 988 AA.
ID 017434
AC 017434
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE MINOR AMPULATE SILK PROTEIN MISPI (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=98200471; PubMed=9541398;
RA Colgin M.A., Lewis R.V.;
RT "Spider minor ampullate silk proteins contain new repetitive sequences
    and highly conserved non-silk-like spacer regions.";
RL Protein Sci. 7:667-672(1998).
DR EMBL: AF027735; AAC14589.1; -.
DR InterPro: IPR000817; -.
DR PRINTS: PR00341; PRION.
FT NON_TER 1
SQ SEQUENCE 988 AA: 79082 MW; 461E03DF53F7085D CRC64;

```

Query Match 45.5%; Score 675; DB 5; Length 988;
 Best Local Similarity 45.4%; Pred. No. 2.4e-39;
 Matches 167; Conservative 6; Mismatches 95; Indels 100; Gaps 12;

```

OY 10 HGSMAGRGLGCGGAG-AAAAAAGAGAGCGGCGY-----GLGSGGTS 53
    |||
DB 240 YRGAGYGGCGGAGAGAGAAAAAAGAGAGAGYGRGAGAGAAAAAGAGAGAGAGY 299
OY 54 GRGGLG-CGGAGAAAAAAG-----GAGCGGCGYGLGSGGT-----90
    |||
DB 300 GCGGCGAGAGAGAAAAAAGAGAGCGGCGYGRGAGAGAAAAAGAGAGAGAGY 359
OY 91 -----SGRGGLGCGGAGAAAAAAGAGAGCGGCGYGLGSGGT-----SGR 131
    |||
DB 360 AAAAAGAGSGGAGYGRGAGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
OY 132 GGLG---CGAGAAAAAAGAGAGAGCGGCGYGLGSGGT-----SGRGGLGCGGAG 177
    |||
DB 420 GGYGRGAGAGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479
OY 178 GAAAAAAGAGAGAGCGGCG-----LGSQTSRGGLG-----CGAGAAAAA 222
    |||
DB 480 GAGAGAGAGAGAGAGAGCGGCGYGRGAGAGAAAAAGAGAGAGAGAGAGAGAGAGAG 539
OY 223 AAAAGAGCGGCGYGLGSGGT-----SGRGGLGCGGAGAAAAAAGAGAGAGAG 271
    |||
DB 540 AGAAAAAGAGGTGGCGGCGYGRGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 597
    |||

```

QY	272	YGCGLSOG	279		1
Db	598	YCGGCGYG	605		
RESULT	4				
ID	Q16988	PRELIMINARY;	PRT;	410	AA.
AC	Q16988;				
DT	01-NOV-1996	(TREMBLREL. 01, Created)			
DT	01-NOV-1996	(TREMBLREL. 01, Last sequence update)			
DT	01-OCT-2000	(TREMBLREL. 15, Last annotation update)			
DE	FIBROIN-4 (FRAGMENT).				
DE	ADF-4.				
GN	Aranneus diadematus (Spider).				
OS	Aranneus diadematus (Spider).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;				
OC	Aranemorphae; Entelegynae; Araneolidae; Araneidae; Araneus.				
OX	NCBI Taxid=45920;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=96178678; PubMed=8600519;				
RA	Guertel P.A., Ginzlinger D.G., Weber B.H., Gosline J.M.;				
RT	"Silk properties determined by gland-specific expression of a spider				
RT	fibroin gene family."				
RT	Science 272:112-115(1996).				
DR	EMBL: U47856; AAC47011.1; -. .				
FT	NON_TER	1			
QO	SEQUENCE	410	AA;	34904	MM; E87A77BD59FF35C CRC64;

Query Match	43.1%	Score 640.5	DB 5	Length 410
Best Local Similarity	56.9%	Pred. No. 2.6e-37		
Matches 169	Conservative 24	Mismatches 61	Indels 43	Gaps 16
Qy	14 ASGRGGLG--GCG-----	AGAAAAAAAAAAGAGCGGCGT	CGTSGRGGTGGCG	62
Dd	11 ASGCGCYPENCGPGPVA	YCGPGVSSAAAAAAGSGCGT	CG--ENCGPSGPGGPGG	69
Qy	63 AGAAAAAAAAAAGCGGCGT	CGTSGRGGCG-----	CGA-----	111
Dd	70 SGSSAAAAAANA--SGPGCYG	CGSGGPGSGCGYCGPGGACGPGG	PGGASAAAAA	126
Qy	112 AGGACGGGCGTGGTSGRGG	CGGAGAAAAAAGAGCGGCGTGGTSGRGG	171	
Dd	127 AASGPGGTCG--GSGCGPGG	PGCATGCGGCGSSAAAAAANA--	SGPGTCG--GSGGPGGTCV	182
Qy	172 LCGGAGAAAAAANAAGAGCGG	CGTGGTSGRGGCGGAGAAAAAAGAGCG	231	
Dd	183 YGPGGPGSSAAAAAANA--	GSCTGCGTGC--ENCGPSGPGG	CGGCGSSAAAAAANA--	237
Qy	232 GGYGGLGSGGTSGRGGTGG	CG-----	GAGAAAAAANAAGAGCGGCGTGGTSG	282
Dd	238 GGYGCG--GSGGPGGPGG	SGGCGSGGSGPGASAAAAAANAAGSGPGTCG	--GSGGPGG	292
RESULT	5			
ID	076786	PRELIMINARY:	PRT: 2639 AA.	
AC	076786:			
DT	01-NOV-1998	(TREMBLREL. 08, Created)		
DT	01-NOV-1998	(TREMBLREL. 08, Last sequence update)		
DT	01-OCT-2000	(TREMBLREL. 15, Last annotation update)		
DE	FIBROIN.			
OS	Antheraea pernyi (Chinese oak silk moth).			
CC	Eukaryota, Metazoa, Arthropoda, Tracheata, Hexapoda, Insecta:			
CC	Pterygota, Neoptera, Endopterygota, Lepidoptera, Glossata, Ditrysia:			
CC	Nymphalidae, Saturniidae, Saturniinae, Antheraea.			
OX	NCB1_TaxID=7119;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Sezutsu H., Tamura T., Yukihiro K.;			
RT	"Characterization of the full length fibroin gene of a wild silkworm, Antheraea pernyi.";			

RL Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF083334; AAC32606.1; -
SQ SEQUENCE 2639 AA; 216056 MW; 2EE3310DEEB09B9A CRC64;

Query Match	41.6%;	Score 617.5;	DB 5;	Length 2639;
Best Local Similarity	39.8%;	Pred. No. 4.9e-35;		
Matches 174;	Conservative 10;	Mismatches 90;	Indels 163;	Gaps 13;

QY	14	ASGAGGGLG-----QGAGAAAAAAAAAGAGGCGYGGGLGSGQT-----	52
Db	1142	ASGAGGSGGGYGMWDGCGYGSDSAAAAAAAAAAAAAAAAASGAGSGYGGYGSDBSAAAAAAAA	1201
QY	53	-----SGRGGLG-----QGAGAAAAAAAAAGAGAQGGYGGGLGSGQT---	90
Db	1202	AAAAAAAAAGSAGAGCGYGMWBDGCGYGSDBSAAAAAAAAAAAAAAAAASGAGSGCGYGGYGSDBSA	1261
QY	91	-----SGRGGLG-----QGAGAAAAAAAA--AGAGCGCGYG-121	
Db	1262	AAAAAAAAAAGSAGAGAGGCGYGMWDGCGYGSDSAAAAAAAAAAAAAAAAAGSAGAGRGDGGYGS	1321
QY	122	-----GLGSGTSGRGGLG-----:11111111	136
Db	1322	GSSAAAAAAAAAAARRAGHDRAAGSAAAAAAAAAAAAAAAAASGAGSGCGYGMWDGCGYSD	1381
QY	137	--QGAGAAAAAAAAAGCAGCGYGGGLGSGQT-----SGRGGLG-----	174
Db	1382	SAAAAAAAAAASGAGSGCGYGGYGSDBSAAAAAAAAAAAAAAAAASGAGYGGYGGYGMWDG	1441
QY	175	-----QGAGAAAAAAAAAGAGCGGCGYGGGLG-----QGSGRGGLG	212
Db	1442	GYGSDSAAAAAAAAAASGAGSGCGYGGYGSDBSAAAAAAAAAAAAAAAAASGAGCGGCTGG	1501
QY	213	QGAGAAAAAAAAAAGAGCGGCGYGGGLGSGTSGRGGLGCGAGAAAAAAAAAAG-gAGCGG	271
Db	1502	YGSDSAAAAAAAAAASGAGAGGCGYGMWDGCGYGSYSAAAAAAAAAAAGSGAGG	1561
QY	272	YGGGLGSGTSGTRRPA	288
Db	1562	DGGYGS-gSSAAAAAAAA	1577

RESULT	6		
ID	002402	PRELIMINARY;	PRT; 738 AA.
AC	002402;		
DT	01-JUL-1997 (TREMBLrel. 04, Created)		
DT	01-JUL-1997 (TREMBLrel. 04, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	INSOLUBLE PROTEIN.		
OS	Pinctada fucata.		
OC	Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidea;		
OC	Pterioidea; Pteriliidae; Pinctada.		
OX	NCBI_TaxID=50426;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=97320490; Pubmed=9177341;		
RA	Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,		
RA	Nakashima K., Takahashi T.;		
RT	"Structures of mollusc shell framework proteins.";		
RL	Nature 387:563-564(1997).		
SR	EMBL; D86074; BAA20466.1; -		
SQ	SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;		

	Query Match	Score	593.5;	DB 5:	Length	738;
	Best Local Similarity	41.1%;	Pred.No.	7.3e-34;		
	Matches	167;	Conservative	10;	Mismatches	82;
					Indels	147;
					Gaps	12;
QY	11 GSWAGCGGGLGGGCGAGAAAAAAGCAGC-----GGYGGLSGQSQTSG---	54				
	:					
Db	113 GGAGCAGGCGGAGGAGGCGACGACAGCAGLGLGLGGLGGLGGLGGLGGGD	172				

Oy	55	-----RCGLGGCGAGAAAAAAGCAGCGGCGTSGTGCTGCQ	99
Db	173	DLFLDLPDDLGAALATLTCGAGGAGCAAAAAAAAAAACGGVGGAAAAAAAGGAGR	222
Oy	100	-GAGAAAAAAAAAAGAG------QGCGTGTGSGTGRGCLGGQC--ACAAAAAAAAA	150
Db	233	LGGAAAAAAGAGAGGAGTGTGCGTGTGCG--GTCGTGTGTGTGCGTGGSSAAAAAAA	291
Oy	151	GGAGCGGCGTGTGSGTGRGGLGCGCAG-----AAAAAAGCAGCGGCGTGTGCG	203
Db	292	AAAGCGGTGCGFYCG--GRCGRRCRGRRRAAAAAAAAAAAGCGGGGGG----	344
Oy	204	TSGHGTGCGCAGCAAAAAA-----	226
Db	345	---GGGGGAGAAAAAASASASAFOMSGTRDALCDIKDLRSNGASAKASAVA	400
Oy	227	-----GAGCGGCGTGTGSGTSGR-----	245
Db	401	STKSQIDLDKVLTKDLAGLLKSSASASASASASAGCGGGGCGGNGGGGGGAGALA	460
Oy	246	-----GGLGGGAGAAAAAAGCAGCGGCGTGTGCGTSG	282
Db	461	AATAAGAGGGLGGGGGAGLAALLAA--GAGGGGFGGLGTGTGCG	505

RESULT	7	
053559		
ID	053559	PRELIMINARY; PRT; 1489 AA.
AC	053559;	
DT	01-JUN-1998 (TREMBLrel. 06, Created)	
DT	01-JUN-1998 (TREMBLrel. 06, last annotation update)	
DT	01-JUN-2000 (TREMBLrel. 14, last annotation update)	
DE	PCRS-FAMILY PROTEIN.	
GN	RV3514 OR MTV023.21.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID:1773;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN:H37Rv;	
RX	MEDLINE:982595987; PubMed:9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,	
RA	Badcock K., Baaham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,	
RA	Hornsbay T., Jajela K., Kroch A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rafterandream M.A., Rogers J.,	
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares J., Sulston J.E.,	
RA	Taylor K., Whitehead S., Barrrell B.G.;	
RT	Deciphering the biology of Mycobacterium tuberculosis from the	
RL	complete genome sequence. ";	
RT	Nature 393:537-544 (1998).	
DR	EMBL: AL022022; CAI17751.1; -.	
DR	Tuberculist; RV3514; -.	
DR	InterPro; IPR000084; -.	
DR	InterPro; IPR002173; -.	
DR	Pfam; PF00934; PE; 1.	
DR	Prodom; PD001223; -; 1.	
DR	PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.	
QO	SEQUENCE 1489 AA; 115754 MW; 6855CBAIC3DBAF3A CRC64;	

Query Match	37.48;	Score 556;	DB 2;	Length 1489;
Best Local Similarity	39.38;	Pred. No. 5e+31;		
Matches 150; Conservative	17;	Mismatches 103;	Indels 112;	Gaps 16

```

0y      11  GSMAGSGGGGCGG_GGAGAAAAAAAAGAGCGGCGGCGG-----GTS 53
          | | | | | | | | | | | | | | | | | | | | | | | |
Db      292  GGALGCGTGTGTGTGGACGACGRCALLLGGAGGGGLGAGGGGGTGGAGGADVLGAGVGTG 351
          | | | | | | | | | | | | | | | | | | | | | | | |
0y      54  GGGGCG-----GGAGAAAAAAAAGGAGCGCGTGCCTGCGTSSRGGLGAGGGA----- 101
          | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	352	GKGGVGGVAGLGGAGGAAAGOLFSSASGAAGNAGVGGAGCGG--	-DPGGAGGAGADADPGAT	409
QY	102	-----GAAAAAAAAA-----	AGGGVGGTSGG-----	SGRGGTGGGGAGA 141
Db	410	GCTGAGGAGGAGGAGGAGSSGCAAGCTGNSGCGAGCGGCGAGGAGAGADNP	TGTGTGGTGGCGTG	469
QY	142	AAAAAAAAAAGAGGCGGCGGCG--	LGSGGTSGRGGTGGGQ-----	GAGAAA-----A 182
Db	470	GAAAGGAGGCGAAGTGTGTGCMGTCTG	GNAGVGGAGCGCGDGGAGGAGADADPGATGCTGFA	529
QY	183	AAAAAAGGA-----	GCGGCGVGLGSGG-----	TSGRGGTGGGGAGAAA 220
Db	530	GAGGAGGAGAGSSGCAAGCTGNSGCGAGCTGGCGGAGAGGAGACADNP	TGTGTGTGGTGGCGTGCA	589
QY	221	AAAAAAAAAGCGGGGCGG--	LGSGGTSGRGGTGGGQ-----	GAGAAAAA-----AAAA 264
Db	590	GAGCGAGCAATGCTGTGCMGTCTG	GNAGVGGAGCGCGDGGAGGAGADADPGATGCTGTFAGGA	649
QY	265	GGAGGG-----	YGLGSGGTSG 282	
Db	650	GGAGTAGGSSSAGGTNSGCSAG	671	

RESULT	8	
053552		
ID	053552	PRELIMINARY; PRT; 1381 AA.
AC	053552;	
DT	01-JUN-1998 (TREMBLrel. 06, Created)	
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)	
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)	
DE	PGS-FAMILY PROTEIN.	
CN	RV3507 OR MTW023.14.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacterinae; Mycobacteriaceae; Mycobacterium.	
OX	[1]	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-H37RV;	
RX	MEDLINE=98295987; PubMed=9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,	
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,	
RA	Hornisy T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Rutter S., Seeger K., Skellern S., Squares S., Squares R., Sulston J.E.,	
RA	Taylor K., Whitehead S., Barrell B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RT	complete genome sequence.";	
RL	Nature 393:537-544(1998).	
DR	EMBL: AL022022; CAAL17744.1; -.	
DR	HSSP: P00778; 2ULL.	
DR	TubercuList: RV3507; -.	
DR	InterPro: IPR000084; -.	
DR	InterPro: IPR002173; -.	
DR	Pfam: PF00934; PI; 1.	
DR	ProDom: PD001223; -; 1.	
DR	PROSITE: PS00583; PFKM_KINASES_1; UNKNOWN_2.	
SO	SEQUENCE 1381 AA; 110624 MW; CA09676BD07F6482 CRC64;	

Query Match	36.7%;	Score 544.5;	DB 2;	Length 1381;
Best Local Similarity	37.8%;	Pred. No. 2.9e-30;		
Matches 144;	Conservative 17;	Mismatches 115;	Indels 105;	Gaps 13;

```

Oy 1 MRSIIHHHHHSMASGSGGGLGCGGAGAAAAAAAAGCAAGCGGGLG -GSGTSGGGL 58
Oy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 MGGTGGCGNGCNCALLTGGGGLG -GAGCMGTGGGTGGTGGCGNCGALLTACGVCAGAGI 262
Oy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 59 GCGCGAAAAAAAAGCAGCGGTGGLGSGTSGRGLGCGGAAAAAAAAGCA --- 115
Oy      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 GCGGTGAGCAAGAGGTG -GAGCGAGGLFPMNG -CDGAGGCGGDBGAAGDAAASAGCGGCK 318

```



```

QY 116 -----GCGYGLGSGQ-GTSGRGGLGCGC-----AG 140
DB 319 GCGCGDGTGTCAGGAGPVLFGHGGAGCGMGCGGTGCMGAGAGDGTIVIAAGTGGEGTGG 378
QY 141 AAAAAAAGAGAGGAGG--YGLGSGOSTSGRGGLGCGGAGAAAAAAG-----GAGCG 194
DB 379 AAGAGGAARGLTSSGLAGVAGAGTGTGTGTGGNGAATAAAYVFGANGDPGAGGKG 438
QY 195 GYGLGSGQ-----GTSGRGGLGSGQ-----GAGAAAAA 221
DB 439 GNGGIGGAANTGVAVAGCGGTGCGKGTGCGAGAGAGNDAGSTGNPGKGGDGGTGCAGAGGA 498
QY 222 AAAAA--AGGAGCGYGLGSGQGTSGRGGLGSGQ-----GAGAAAAA 261
DB 499 AGTNGGAGHAAGTDCGGCGGTGCGNGCGTGVNGADNTLNPDPFGAGEPGAGAGAGAGG 558
QY 262 AAGGAGGCGYGLGSGQGTSG 282
DB 559 AAGGPGGTGTGGNGGNG 579

RESULT 9
ID 053557 PRELIMINARY; PRT: 1079 AA.
AC 053557;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PGRS-FAMILY PROTEIN (FRAGMENT).
GN RV3512 OR MTV023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornbly T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL022022; CAI17749.1; -.
DR Tuberculist; RV3512; -.
DR InterPro: IPR002202; -.
DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
FT NON_TER 1 1
SQ SEQUENCE 1079 AA; 81163 MW; A79718CDBC874B97D CRC64;

Query Match 36.6%; Score 543.5; DB 2; Length 1079;
Best Local Similarity 40.7%; Pred. No. 2.7e-30;
Matches 144; Conservative 12; Mismatches 109; Indels 89; Gaps 14;

```

```

QY 149 AAGAG-----GCGYGLGSGQGTSGRGGLGCGGAGAAAAAAGAGCGYGL-- 199
DB 395 GAGGAGCGSGPNTSPGCGGCGGCGGCGAG-GAAGAGAGGAGGAGTGANCGGCGAGTGG 453
QY 200 -----GSGTSGRGGLGCGGAGAAAAAAGAGCGYGLGSGQGTSGRG 246
DB 454 AGAASATNGSGGAGGTGGGSGGAGGTGAGAGTGGAAAGGGGCGGAGG-GAGGCG 512
QY 247 GLGSGQ-----AGAAAAAAGAGAG-GGCGY-GGLGSGQGTSG 282
DB 513 GAGAGCGTGCNGNITGTAGTAAAGNCGAAGKAGAGCGCGGTGTGGGAG 566

RESULT 10
ID 016986 PRELIMINARY; PRT: 294 AA.
AC 016986;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE FIBROIN-2 (FRAGMENT).
GN ADF-2.
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
NCBI_TaxID=45920;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96178678; PubMed=8600519;
RA Guerette P.A., Ginzinger D.G., Weber B.H., Gosline J.M.;
RT Silk properties determined by gland-specific expression of a spider
RT fibroin gene family."
RL Science 272:112-115(1996).
DR EMBL: U47854; AAC47009.1; -.
FT NON_TER 1 1
SQ SEQUENCE 294 AA; 24729 MW; BBCTE18AE2C9628 CRC64;

Query Match 36.5%; Score 541.5; DB 5; Length 294;
Best Local Similarity 52.3%; Pred. No. 1.2e-30;
Matches 136; Conservative 9; Mismatches 46; Indels 69; Gaps 10;

```

OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RA EMBL; AL021841; CAA17117.1; -.
DR HSSP; P00441; 1SOS.
DR Tuberculist; RV3345c; -.
DR InterPro; IPR000084; -.
DR InterPro; IPR002173; -.
DR Pfam; PF00934; pf.1.
DR Prodom; PD001223; -1.
DR PROSITE; PS00583; PKB_KINASES_1; UNKNOWN_3.
SQ SEQUENCE 1538 AA; 129386 MW; 788F0B2095587592 CRC64;

Query Match 36.2%; Score 538; DB 2; Length 1538;
Best Local Similarity 38.9%; Pred. No. 8.8e-30;
Matches 126; Conservative 13; Mismatches 131; Indels 54; Gaps 9;

QY 11 GSMASRGGLGCGAGAAAAAAGAGAGCGGCGTSGRGLGCGAGAAAAA 70
DB 1030 GMINGLGFGAGGAGGAVDAATTGCAAGNGAGGFASTGIGRGAGCGPAGADFPAG 1089
QY 71 AAAAGCA---GQGGYGGIGLSG-----TSGRGLGCGAGAAAAAAGG----- 114
DB 1090 VCGVGAGGDDGCGGCGGIGGEGRTGNGSGGDDGGGIGLGGNGLGGNGGVS 1149
QY 115 -----AGGGYGGIGLSGCTSGRGLGCGAGAAAAAAGAGCGGCGTSG----- 164
DB 1150 ETGFGAGGNGGYGGP--GPEGNGLGGNGAGGNGVSTTGGDGGAGGNGGDDGN 1207
QY 165 -----GTSGRGLGCGAGAAAAAAGG-----AGCGYGGIGLSGCTSG 205
DB 1208 VGLGGDAGSGGAGNGGCTGTGAGAGAGAGAGNGSSSKSTTTGNAGSGAGNGGTGIAN 1267
QY 206 GRGGLGCGAGAAAAA---AAAGGAGCGGYGIGLSGCTSG--GCGGLGCGAGAAAAA 261
DB 1268 GAGGAGGAGGAGNAGVAGVAGCGDGGNGGNGHGGDGTGAGAGGAGGNGSSGAASGSG 1327
QY 262 ---AAAGGAGCGGYGIGLSGCTSG 282
DB 1328 VVNTVAGHGGNGGNGGNGGNSAG 1351

RESULT 12
OS3844 PRELIMINARY; PRT; 749 AA.
AC 053844;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PGRS-FAMILY PROTEIN.
GN RV0833 OR MTV043.25.

OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE 749 AA; 57448 MW; C3C91307F9A8A2BC CRC64;

Query Match 36.0%; Score 535; DB 2; Length 749;
Best Local Similarity 39.9%; Pred. No. 7.7e-30;
Matches 133; Conservative 9; Mismatches 123; Indels 68; Gaps 11;

QY 16 GRGGLGCGAGAAAAAAGAGAGCGGCGTSGRGLGCGAGAAAAA 55
DB 3 GNGAGGSSGAPGATGAGGPAAGLIGVCGAGAGGDSNAGVTCGAGAGGAAALFGAGGA 62
QY 56 GGLGCGAGAAAAAAGCA-----GQGGYGGIGLSGCTSGRGLGCGAGAAAAA 108
DB 63 GAGGSSG--GSGAAGGAGAGAGAGGLFASGSGSGFCGFASTGTGAGAGTGAAGLFPASGV 121
QY 109 AAAAGAGCGGYGIGLSG-----TSGRGLGCGAGAAAAAAGAGCGGCGTSG 161
DB 122 GGTGGAGSGGTGAGGAGLFPASGAGAGSGGTGTGAGGTGAGAGLFPAGAGAGL 181
QY 162 GSQG--TSGRGLGCG-----GAGAAAAAAGAGCA--GQGGYGL--GSQGTSGR 207
DB 182 GCGNHTGCHGAGGSAGLALGDDGAGAGGAAATTGTGAGAGAGKAGLFLGSGGAGGS 241
QY 208 GGLGCG-----GAGAAAAAAGAGAGCGGYGIGLSGCTSGRGLGCGAGCA---- 255
DB 242 GGAAGTFCDTGNSGAGAGAGKAGLLFSGCAGSGSAGGFANCGSTGAGAGAGGAGLIG 301
QY 256 -----AAAAAAGAGAGCGGYGIGLSGCTSGI 283
DB 302 NGNGSGGTGATGAGAGGAGAG--GGAAGL 332

RESULT 13
OS3119 PRELIMINARY; PRT; 421 AA.
AC 093119;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ANTHEERA PERNYI FIBROIN (FRAGMENT).
OS Antheraea pernyi (Chinese oak silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Saturniidae; Saturniinae; Antheraea.
OX NCBI_TaxID=7119;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POSTERIOR SILKGANDS;
RA Yukuhiro K., Kanda T., Tamura T.;
RT "Preferential codon usage and two types of repetitive motifs in the
RT fibroin gene of the Chinese oak silkworm, Antheraea pernyi.";
RN Insect Mol. Biol. 0:0-0(1996).

XX The present sequence represents the natural spider silk protein
CC spidroine major 1. The protein improves the moisturizing/softening
CC action of the compositions. The protein, and its fragments are used
CC in cosmetic or dermatological compositions. These compositions have
CC use as hair or skin care products; and make-up or sunscreens.
CC As the protein is a good, persistent film-formers on the skin
CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.

SQ Sequence 651 AA:

Query Match 61.1%; Score 2135; DB 20; Length 651;
Best Local Similarity 68.3%; Pred. No. 1.4e-148;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

```
23 OGAGAAAAAAGAGAGGCGGCGGCGT--SGRGLGCGAG-AAAAAAAAAGAGAG 79
||||| ||||||| ||||||| || ||||||| ||||||| ||||||| |||||||
1 ggaq-----aaaaaagagaggggylgggagggggylgggagaggaagaaagagag 56
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
80 GGTGGLSGGTSGRGLGCGAGAAAAAAGAGAGCGGTGCGTSGRGLGCGAG 139
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
57 ggggylgggq-----agrgggag-----aaaaaagagggggylgggag- 107
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
140 GAAAAAAGAGAGAGCGGCGTSGPGGPGGQGTSGRGLGCGAGAAAAA 199
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
108 g-----aaaaaagagggggylgggq-----agrggg-----aaaaa 142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
200 AGAGAGCGGTGCGTSGRGLGCGAGAAAAA-----AGAGCGGTGCG 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
143 aggaagggggylgggq-aggylgggagaaaaaagagggggylgggagggggylg 201
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 SGTSGRGLGCGAGAAAAA-----AAAGAGAGCGGTGCG 289
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
202 sgg-aggylgggagaaaaaagagggylgggagggagaaagagggggylgg 260
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
290 GGTSGRGLGCGAGAAAAAAGAGAGCGGTGCGTSGPGGPG-GOGTSGRGL 348
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
261 gq-----agrgggag-----aaaaaagagggggylgggq-aggggggylgg 311
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
349 GCGAGAAAA-----AAAAAGAGAGCGGTGCGTSGRGLGCG 390
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
312 ggggagaaaaagagggylgggagggagaaaaaagagggggylggg-agg 370
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
391 AGAAAAAAGAGAGCGGTGCGTSGRGLGCGAGAAAAAAGAGAGCGGTG 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
371 ag-----avaaaagagggggylgggq-----agrgggag-----aaaaa 419
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
451 LGSGTSGRGLGCGAGAAAAAAGAGAGCGGTGCGTSGRGLGCGT-----SGP 497
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
420 lgnng-aggylgggag-----aaaaaagagggggylgggagggagaaagag 475
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
498 GCGGP-GOGTSGRGLGCGAGAAAAA-----AGAGCGGTGCGTSGRGL 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
476 ggggylgggag-----gggagaaaaaagagggylgggagggggylgg 531
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
551 GCGAGAAAA-----AAAAAGAGAGCGGTGCGTSGRGLGCG 590
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
532 ggggagaaaaagagggylgggagggagaaaaaagggggylgggq-----ag 587
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
591 OGAGAAAAAAGAGAGCGGTGCGTSGRGLGCGAGAAAAAAGAGAGCGGT 650
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
568 ggaq-----aaaaaagagggggylgggq-v-grggylgggag-----aaaa 636
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
651 GGLGS 655
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
637 gggg 641
```

RESULT 2

AAW53346
ID AAW53346 standard; Protein; 718 AA.

AC AAW53346;

DT 06-JUL-1998 (first entry)

DE Nephila clavipes spider silk protein.

XX Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;
XX cocoon; tensile strength; elasticity.

XX Nephila clavipes.

XX OS Nephila clavipes.

XX PN US5728810-A.

XX PD 17-MAR-1998.

XX PF 19-APR-1995; 95US-0425069.

XX PR 15-APR-1991; 91US-0684819.

XX PR 20-APR-1990; 90US-0511792.

XX PR 04-OCT-1994; 94US-0317844.

XX PR 19-APR-1995; 95US-0425069.

XX PA (UTWT-) UNIV WYOMING.

XX PI Human MB, Lewis RV, Xu M;

XX PS WPI: 1998-270437/24.

XX DR N-PSDB; AAV23249.

XX PS Recombinant spider silk proteins - useful for making fibres

XX PS Claim 1; Column 29-34; 68pp; English.

XX The present sequence represents a spider silk protein from the present
CC invention. Spider silk proteins, and peptide fragments within the
CC proteins, can be produced and purified independently and can then be
CC mixed and made into fibres that have higher tensile strengths and
CC elasticity than naturally occurring fibres. The fibres can be used in
CC mixed composites. The invention allows the two naturally occurring
CC Nephila clavipes silk proteins to be produced independently so that
CC they can later be combined to form silk fibres of high tensile strength
CC and elasticity.

SQ Sequence 718 AA:

Query Match 61.1%; Score 2135; DB 19; Length 718;
Best Local Similarity 68.3%; Pred. No. 1.5e-148;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

```
23 OGAGAAAAAAGAGAGCGGCGGCGT--SGRGLGCGAG-AAAAAAAAAGAGAG 79
||||| ||||||| ||||||| || ||||||| ||||||| ||||||| |||||||
1 ggaq-----aaaaaagagggggylgggagggggylgggagggagaaagagag 56
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
80 GGTGGLSGGTSGRGLGCGAGAAAAAAGAGAGCGGTGCGTSGRGLGCGAG 139
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
57 ggggylgggq-----agrgggag-----aaaaaagagggggylgggag-agg 107
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
140 GAAAAAAGAGAGAGCGGTGCGTSGPGGPGGQGTSGRGLGCGAGAAAAA 199
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
108 g-----aaaaaagagggggylgggq-----agrggg-----aaaaa 142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
200 AGAGAGCGGTGCGTSGRGLGCGAGAAAAA-----AGAGCGGTGCG 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
143 aggaagggggylgggq-aggylgggagaaaaaagagggggylgggagggggylg 201
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 SGTSGRGLGCGAGAAAAA-----AAAGAGAGCGGTGCG 289
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
202 sgg-aggylgggagaaaaaagagggylgggagggagaaaaaagagggggylgg 260
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

[illegible]

CC	RESULT	3	
CC	AAV59070		
CC	ID	AAV59070	standard; Protein; 718 AA.
CC	AC		
CC	XX	AAV59070;	
CC	XX		
CC	DT	08-MAR-2000	(first entry)
CC	XX		
CC	DE	N. clavipes	spider silk protein 1.
CC	XX		
CC	KM	Spider silk protein;	dragline silk protein; major ampullate gland; fiber
CC	XX		
CC	OS	Nephila clavipes.	
CC	XX		
CC	PN	US5989894-A.	
CC	XX		
CC	PD	23-NOV-1999.	
CC	XX		
CC	PF	04-OCT-1994;	94US-0317844.
CC	XX		
CC	PR	15-APR-1991;	91US-0684819.
CC	XX	20-APR-1990;	90US-0511792.
CC	XX		
CC	PA	(UYWY-) UNIV WYOMING.	
CC	XX		
CC	PI	Hinman MB, Xu M, Lewis RV;	
CC	DR	WPI: 2000-061225/05.	
CC	DR	N-PSDB: AA238195.	
CC	XX		
CC	PT	Isolated DNA, vector and transformed cell	encoding for and useful in
CC	XX	the production of spider silk protein	
CC	PS		
CC	XX	claim 1: Fig 6A-D: 65pp; English.	

CC production of fibers. The present sequence represents the spider silk
CC protein 1, derived from the major ampullate gland of *Nephila clavipes*
XX
50 Sequence 718 AA;

Query Match	61.1%;	Score 2135;	DB 21;	Length 718;
Best Local Similarity	68.3%;	Pred. No. 1.5e-148;		
Matches 495;	Conservative 15;	Mismatches 39;	Indels 176;	Gaps 34;

[illegible]

KM protein superfibre; major ampullate silk; orb web spider.
 XX
 OS *Nephilia clavipes*.
 PN EP452925-A.
 XX
 PD 23-OCT-1991.
 XX
 PF 18-APR-1991; 91EP-0106217.
 XX
 PR 20-APR-1990; 90US-0511792.
 XX
 PA (UYWY-) UNIV OF WYOMING.
 XX
 PI Lewis RV, Xu M, Hinman M;
 XX WPI: 1991-312199/43.
 DR N-PSDB; AAQ14183.
 XX
 PT DNA encoding spider silk protein-1 and 2 and variants - isolated
 PT from *Nephila clavipes*, for prodn. of spider silk protein and
 PT fibres having desired characteristics
 XX
 PS Claim 15; Page 23; 48pp; English.
 XX
 XX The spider silk protein contains a basic 34 amino acid repeat. The
 CC repeat itself contains 3 regions. The first comprises 0-9 amino
 CC acids with a sequence AGR(GGX)2. This region is not highly
 CC conserved. The second region has a sequence GAG(AIX which is highly
 CC conserved and is 8-10 amino acids long. The third segment is (GSX)5
 CC and is 15 amino acids long and is very highly conserved. In most
 CC cases X is A/Q, Y or L. Removal of the poly-(Ala) segments results
 CC in a silk having lower elasticity.

... Sequence 718 AA;

Query Match	60.9%	Score 2126	DB 12	Length 718
Best Local Similarity	68.4%	Pred. No. 6.7e-148		
Matches 492: Conservative	16	Mismatches 47	Indels 164	Gaps 34

[illegible][illegible]

RESULT 5
AAR99053

AC AAR99053;

DT 17-JAN-1997 (first entry)

Spider dragline variant, DP-1A.9 polymer.

Spider; dragline protein; variant; monomer; polymer;
KW

KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;

rope; surgical suture; implant; reinforcement; film; coating.

AA
OS Synthetic.

AA
PN W09429450-A2.

XX
PD 22-DEC-1994

XX
PF 15-JUN-1994:

XX 15-JUN-1993: 93US-0077600.
PR

XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.XX
PT Fahnestock SR:XX
XX
WPT: 1995-036479/05.

XX New synthetic variants of spider dragline protein - for making
PT

etc. also related DNA, vectors and transformed cells

Disclosure: Page 85-87: 168pp: English.
XX
PS

XX This sequence represents a synthetic spider dragline variant polymer, CC DP-1A.9. The sequence of the DP-1A.9 monomer is given in AA899052. CC The polypeptide monomer is a variant based on a consensus sequence CC derived from the fibre forming regions of spider dragline protein, CC esp. the natural protein 1 (spidroin 1) from *Nephila clavipes*. CC DNA sequence encoding the monomer may be used in the recombinant CC production of the variant protein in a recombinant host, e.g. *E. coli* CC or *Bacillus subtilis*. Synthetic analogues of DP1 were designed to mimic CC the repeating consensus sequence of the natural protein and the pattern CC of variation among individual repeats. DP-1A analogues are composed CC of a tandemly repeated 101 amino acid monomer which comprises four CC repeats which differ from the consensus sequence given in AA006201, CC according to the pattern (1)-(5) given below. This 101 amino acid CC monomer is repeated 1-16 times in a series of analogue proteins. The CC individual repeats differ from the consensus according to the pattern:

(1) the poly-alanine sequence varies in length from 0-7 CC residues; (2) when the entire poly-alanine sequence is deleted, CC so also is the surrounding sequence encompassing AGKGGAGGACAGNGG; CC (3) aside from the poly-alanine sequence, deletions usually

CC encompass integral multiples of three consecutive residues:
 CC (4) deletion of GYG is generally accompanied by deletion of GRC
 CC in the same sequence; and
 CC (5) a repeat in which the entire poly-alanine sequence is
 CC deleted is generally preceded by a repeat containing six alanine
 CC residues.
 CC The proteins may be used to produce fibres of high tensile strength and
 CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
 CC for implants, plastic reinforcements, films, coatings, etc.
 CC
 SQ Sequence 606 AA:

Query Match 59.4%; Score 2076; DB 16; Length 606;
 Best Local Similarity 69.7%; Pred. No. 2.5e-144;
 Matches 477; Conservative 11; Mismatches 74; Indels 122; Gaps 30;

OY 19 GLGGGAGAAAAAAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAGAG 78
 DB 3 grrgggag-----aaaaagagggggygglgsgg-agrvglgggag-----aaaaagag 54
 OY 79 QGCGYGLGSGCTSGRCGLGCGGAGAAAAAAGAGGCGGCGTSGRCGLG 136
 DB 55 q-----gglgsgagagagaaaaagagggggygglgsggagggggygglg 100
 OY 137 QGA---GAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAG 185
 DB 101 qgggagggagaaaaagagggggygglgsggagrgglgsggagaaaaagaggg 160
 OY 186 GCGAGAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAG 240
 DB 161 aggagggagaaaaagagggggygglgsggagggggygglgsggagaaaaag 220
 OY 241 AGCGYGLGSGCTSGRCGLGCGGAGAAAAAAGAGGCGGCGTSGRCGLG 300
 DB 221 agggggygglgsgg-agrvglgggag-----aaaaagagag-----gglg 262
 OY 301 QGAGAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAG 359
 DB 263 qgggagggagaaaaagagggggygglgsgg-aggggygglgsggag-----ggggag 314
 OY 360 AAAAAAGAGAGGCGTSGRCGLGCGGAGAAAAAAGAGGCGGCGTSG 419
 DB 315 aaaaaagaggggygglgsgg-agrvglgggag-----aaaaagagag----- 358
 OY 420 GRCGLGCGGAGAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAA 474
 DB 359 --gglgsgagagagaaaaagagggggygglgsggagggggygglgsggagaa 416
 OY 475 AAAAAAGAGGCGTSGRCGLGCGGAGAAAAAAGAGGCGGCGTSG 526
 DB 417 aaaaaagaggggygglgsggagrgglgsggagaaaaagagggggygglgsggag 476
 OY 527 AAGCAGCGCGTSGRCGLGCGGAGAAAAAAGAGAGGCGGCGTSG 581
 DB 477 aagagggggygglgsggagggggygglgsggagrggagagaaaaagaggggygglgsgg 536
 OY 582 TSGRCGLGCGGAGAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAA 641
 DB 537 -agrgglgsggag-----aaaaagagag-----gglgsgagagagagaaaa 578
 OY 642 AGGAGCGGCGTSGRCGLGCGGAGAAAAAAGAGAGGCGGCGTSG 665
 DB 579 aggaggggygglgsgg-aggggygglgsgg 601

RESULT 6
 AAT40100
 ID AAT40100 standard; protein; 606 AA.
 AC AAT40100;
 XX
 DT 19-NOV-1999 (first entry)

XX Polymer of an analogue of spider silk protein spidroine major 1.
 DE Spider silk protein; spidroine major 1; cosmetic; make-up;
 XX dermatological compositions; hair care; skin care; sunscreen;
 KW hormone; moisturizer; skin disorder; skin disorder.
 XX Synthetic.
 OS Nephila clavipes.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..101
 FT /note="monomer unit"
 XX
 PN FR2774588-A1.
 PD 13-AUG-1999.
 XX
 PF 11-FEB-1998; 98FR-0001614.
 XX
 PR 11-FEB-1998; 98FR-0001614.
 XX
 PA (OREA) L'OREAL SA.
 XX
 PI Philippe M, Garson JC, Arraudeau JP;
 PT WPI; 1999-510729/43.
 PT Cosmetic or dermatological composition containing spider silk protein,
 PT for hair or skin care, in make-up or sunscreens
 PS Claim 8; Fig 4B; 32pp; French.
 XX
 CC The present sequence represents a polymer of an analogue of the spider
 CC silk protein spidroine major 1. The protein improves the moisturizing/
 CC softening action of the compositions. The protein, and its fragments
 CC are used in cosmetic or dermatological compositions. These compositions
 CC have use as hair or skin care products; and make-up or sunscreens.
 CC As the protein is a good, persistent film-formers on the skin
 CC of low surface density, it can be used for delivery of active
 CC agents that are generally difficult to administer, e.g. vitamins,
 CC hormones, moisturizers or agents for treating disorders of the
 CC skin and hair.
 CC
 SQ Sequence 606 AA:

Query Match 59.4%; Score 2076; DB 20; Length 606;
 Best Local Similarity 69.7%; Pred. No. 2.5e-144;
 Matches 477; Conservative 11; Mismatches 74; Indels 122; Gaps 30;

OY 19 GLGGGAGAAAAAAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAGAG 78
 DB 3 grrgggag-----aaaaagagggggygglgsgg-agrvglgggag-----aaaaagag 54
 OY 79 QGCGYGLGSGCTSGRCGLGCGGAGAAAAAAGAGGCGGCGTSGRCGLG 136
 DB 55 q-----gglgsgagagagaaaaagagggggygglgsggagggggygglg 100
 OY 137 QGA---GAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAG 185
 DB 101 qgggagggagaaaaagagggggygglgsggagrgglgsggagaaaaagaggg 160
 OY 186 GCGAGAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAG 240
 DB 161 sgagagggagaaaaagagggggygglgsggagrggagrgggagagaaaaag 220
 OY 241 AGCGYGLGSGCTSGRCGLGCGGAGAAAAAAGAGAGGCGGCGTSGRCGLG 300
 DB 221 aggggygglgsgg-agrvglgggag-----aaaaagagag-----gglg 262
 OY 301 QGAGAAAAAAGAGAGGCGGCGTSGRCGLGCGGAGAAAAAAGAG 359
 DB 315 aaaaaagaggggygglgsgg-agrvglgggag-----aaaaagagag----- 358


```
OY 501 GPCQOTSGRGGLGAGAGAAAAAGAGGCGYGLGSGCTSGRG---GLGGGAG 556
Db 535 gP9q9p9y9p9q9p9spsaaagp9y9p-9q9p9y9p9q9p9y9p9 593
OY 557 ---AAAAAGAGGCGYGLGSGCTSGRGGLG-GGAGAGAAAAAGAGGCG 611
Db 594 q9p9p9saaagp9q9p9y9p-9q9p9y9p9q9p9spsaaagp9 652
OY 612 YG-GLGSGCTSGRGGLGAGAGAAAAAG---GAGGCGYGLGSGCTSGPGYGP 666
Db 653 y9p9q9p9y9p9q9p9spsaaagp9y9p9p9q9p9y9p-9q9p9y9p 711
OY 667 GQO 669
Db 712 9q9 714
```

Search completed: July 3, 2001, 14:58:08
Job time: 438 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:27 ; Search time 48.42 Seconds

(Without alignments)
283.323 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493
Sequence: 1 MRGSHHHHSGMSAGRGCL.....GGCGGQQTSGIRPAKLIN 681Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2135	61.1	718	1	US-08-425-069-2 Sequence 2, Appl1
2	2135	61.1	718	2	US-08-317-844B-2 Sequence 2, Appl1
3	2135	61.1	747	4	US-09-034-177-3 Sequence 2, Appl1
4	1405	40.2	832	1	US-08-209-747-2 Sequence 2, Appl1
5	1405	40.2	832	1	US-08-458-298-2 Sequence 2, Appl1
6	1233.5	35.3	1177	1	US-07-609-716-31 Sequence 31, Appl1
7	1233.5	35.3	1177	1	US-08-175-155-29 Sequence 29, Appl1
8	1233.5	35.3	1177	1	US-08-477-509B-64 Sequence 64, Appl1
9	1233.5	35.3	1177	2	US-08-707-237A-35 Sequence 35, Appl1
10	1233.5	35.3	1177	3	US-08-482-085B-64 Sequence 64, Appl1
11	1233.5	35.3	1177	4	US-08-475-411A-31 Sequence 31, Appl1
12	1233.5	35.3	1177	4	US-08-478-029A-31 Sequence 31, Appl1
13	1176.5	33.7	1059	1	US-08-175-155-48 Sequence 48, Appl1
14	1176.5	33.7	1059	2	US-08-707-237A-54 Sequence 54, Appl1
15	1176.5	33.7	1101	1	US-08-477-509B-83 Sequence 83, Appl1
16	1176.5	33.7	1101	3	US-08-482-085B-83 Sequence 83, Appl1
17	1157	33.1	738	3	US-08-664-038A-3 Sequence 3, Appl1
18	1157	33.1	738	3	US-08-664-038A-3 Sequence 3, Appl1
19	1132	32.4	595	2	US-08-425-069-4 Sequence 4, Appl1
20	1132	32.4	595	2	US-07-609-716-41 Sequence 41, Appl1
21	1131	32.4	1332	4	US-08-475-411A-41 Sequence 41, Appl1
22	1131	32.4	1332	4	US-08-478-029A-41 Sequence 41, Appl1
23	1129	32.3	1038	4	US-07-609-716-36 Sequence 36, Appl1
24	1129	32.3	1038	4	US-08-475-411A-36 Sequence 36, Appl1
25	1129	32.3	1038	4	US-08-478-029A-36 Sequence 36, Appl1
26	1127	32.3	979	1	US-08-477-509B-89 Sequence 89, Appl1
27	1127	32.3	979	3	US-08-482-085B-89 Sequence 89, Appl1

28	1127	32.3	1050	1	US-08-175-155-54 Sequence 54, Appl1
29	1103.5	31.6	1018	1	US-08-089-862-11 Sequence 11, Appl1
30	1103.5	31.6	1018	1	US-08-587-333-18 Sequence 18, Appl1
31	1103.5	31.6	1018	5	PCT-US94-07776-16 Sequence 4, Appl1
32	1099	31.5	907	2	US-09-010-928B-4 Sequence 6, Appl1
33	1095.5	31.4	945	1	US-08-089-862-6 Sequence 13, Appl1
34	1095.5	31.4	945	1	US-08-587-333-13 Sequence 11, Appl1
35	1095.5	31.4	945	5	PCT-US94-07776-11 Sequence 80, Appl1
36	1094	31.3	2100	1	US-08-477-509B-80 Sequence 80, Appl1
37	1094	31.3	2100	3	US-08-482-085B-80 Sequence 80, Appl1
38	1094	31.3	2107	1	US-08-175-155-45 Sequence 51, Appl1
39	1094	31.3	2107	2	US-08-707-237A-51 Sequence 28, Appl1
40	1083	31.0	745	2	US-09-010-928B-28 Sequence 53, Appl1
41	1082	31.0	766	1	US-08-175-155-53 Sequence 88, Appl1
42	1082	31.0	766	2	US-08-477-509B-88 Sequence 61, Appl1
43	1082	31.0	766	3	US-08-707-237A-61 Sequence 88, Appl1
44	1082	31.0	766	3	US-08-482-085B-88 Sequence 88, Appl1
45	1080.5	30.9	870	2	US-09-010-928B-2 Sequence 2, Appl1

ALIGNMENTS

RESULT 1
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himman, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match 61.1%; Score 2135; DB 1; Length 718;
Best Local Similarity 68.3%; Pred. No. 7.4e-151;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;
OY 23 OGAGAAAAAGAGAGCGGTGGTGGT--SGRGGLGGGAG-AAAAAAAAAGAGAG 79

COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175.155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Berttram I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-175-155-29

Query Match 35.3%; Score 1233.5; DB 1; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3.5e-84;

Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps 11;

QY 13 MASGRGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---GGAGAAAAA 69
DB 32 MGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 91
QY 70 AAAAAAGAGCGGCGGCGTSGRGLGGGAGAAAAAAGAGAGAGCGGCGTSGCT 128
DB 92 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 151
QY 129 SGRGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLGSGTSGPGYGPQQT---SGRGLG 185
DB 152 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 211
QY 186 GCGAGAAAAAAGAGAGGCGGCGGCGTSGRGLGSGTSGRGLG---GLGCGAGAAAAAAGAGAG 242
DB 212 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 271
QY 243 GCGGCGGCGTSGRGLGGGAGAAAAAAGAGAGAGCGGCGG---GLGSGCTSGR 295
DB 272 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 331
QY 296 GGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLGSGTSGPGYGPQQT---SGRGLGGG 352
DB 332 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 391
QY 353 AGAAAAAAGAGAGGCGGCGGCGTSGRGLG---GLGCGAGAAAAAAGAGAGAGGCGG 409
DB 392 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 451
QY 410 YGGLGSGCTSGRGLGGGAGAAAAAAGAGAGAGCGGCGG---GLGSGCTSGRGL 462
DB 452 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 511
QY 463 GGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLGSGTSGPGYGPQQT---SGRGLGGGAG 519
DB 512 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 571
QY 530 AAAAAAAGAGAGGCGGCGGCGTSGRGLG---GLGCGAGAAAAAAGAGAGAGGCGG 576
DB 572 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 631
QY 577 LGSQGTSGRGLGGGAGAAAAAAGAGAGAGCGGCGG---GLGSGCTSGRGLGG 629

DB 632 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 691
QY 630 GAGAAAAAAGAGAGGCGGCGGCGTSGRGLGSGTSGPGYGPQ 667
DB 692 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 729

RESULT 8
US-08-477-509B-64

Sequence 64, Application US/08477509B

Patent No. 5770697

GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A

APPLICANT: Crissman, John W

APPLICANT: Dorman, Mary A

TITLE OF INVENTION: No. 5770697e1 Peptides Comprising Repetitive

TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same

NUMBER OF SEQUENCES: 112

CORRESPONDENCE ADDRESS:

ADDRESS: Flehr, Hobach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: US

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentlin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477.509B

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/175.155

FILING DATE: 29-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/053.049

FILING DATE: 22-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/114.618

FILING DATE: 29-OCT-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 06/927.258

FILING DATE: 04-NOV-1986

ATTORNEY/AGENT INFORMATION:

NAME: Treccarlin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-55186-7/RTV/MTK

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989

TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 1177 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-477-509B-64

Query Match 35.3%; Score 1233.5; DB 1; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3.5e-84;

Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps 11;

QY 13 MASGRGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---GGAGAAAAA 69
DB 32 MGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 91


```

Oy 70 AAAAAAGCGCGCGTGLGSGCTSGGGGGLGCGGAGAAAAAAAAG--GAGCGGCGGLGSGGT
Db 92 GAGSGAGAGSAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG
Oy 129 SGRGGLGCGGAGAAAAAAAAGAGAGCGGCGTGLGSGCTSGPGCGPGGQOT---SGRGGLG
Db 152 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA
Oy 186 GCGGAGAAAAAAAAGAGCGCGTGLGSGCTSGRG---GLGCGGAGAAAAAAAAGAG
Db 212 GSGGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG
Oy 243 GCGGCGGSGCTSGRGGLGCGGAGAAAAAAAAGAGCGGCGT-----GLGSGGTSGR
Db 272 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG
Oy 296 GGLGCGCGGAGAAAAAAAAGAGCGGCGTGLGSGCTSGPGCGPGGQOT---SGRGGLGCGG
Db 332 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG
Oy 353 AGAAAAAAAAGAGAGCGCGTGLGSGCTSGRG---GLGCGGAGAAAAAAAAGAGCGG
Db 392 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA
Oy 410 YCGGSGCTSGRGGLGCGGAGAAAAAAAAGAGCGGCGT-----GLGSGGTSGRGGL
Db 452 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG
Oy 463 GCGGAGAGAAAAAAAAGAGCGCGTGLGSGCTSGPGCGPGGQOT---SGRGGLGCGGAG
Db 512 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA
Oy 520 AAAAAAAGAGAGCGGCGTGLGSGCTSGRG---GLGCGGAGAAAAAAAAGAGCGGCGT
Db 572 GSGGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG
Oy 577 LGSGCTSGRGGLGCGGAGAAAAAAAAGAGCGGCGT-----GLGSGGTSGRGGLGCG
Db 632 SGAAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG
Oy 630 GAGAGAGAAAAAAGAGAGCGCGTGLGSGCTSGPGCGPGG 667
Db 692 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 729

RESULT 9
US-08-707-237A-35
Sequence 35, Application US/08707237A
Patent No. 5830713
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Capello, Joseph
APPLICANT: Crisman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
NUMBER OF INVENTION: REPETITIVE DNA
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/707,237A
FILING DATE: 03-SEP-1996
CLASSIFICATION: 435

```

```

1  PRIOR APPLICATION DATA:
2  APPLICATION NUMBER: US 08/175,155
3  FILING DATE: 29-DEC-1993
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER: US 08/053,049
6  FILING DATE: 22-APR-1993
7  PRIOR APPLICATION DATA:
8  APPLICATION NUMBER: US 07/609,716
9  FILING DATE: 06-NOV-1990
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 07/269,429
12 FILING DATE: 09-NOV-1988
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: US 07/114,618
15 FILING DATE: 29-OCT-1987
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 06/927,258
18 FILING DATE: 04-NOV-1986
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Treccartin, Richard F.
21 REGISTRATION NUMBER: 31,801
22 REFERENCE/DOCKET NUMBER: A-55186-10/WHD
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (415) 781-1989
25 TELEFAX: (415) 398-3249
26 TEXT: 910 277299
27 INFORMATION FOR SEQ ID NO: 35:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 1177 amino acids
30 TYPE: amino acid
31 STRANDEDNESS: unknown
32 TOPOLOGY: unknown
33 MOLECULE TYPE: protein
34 US-08-707-237A-35
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956

```

[illegible]

```

US-08-482--085B-64

Query Match          35.3%; Score 1233.5; DB 3; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3.5e-84;
Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps

QY      13 MASRGGTGTGCGACAAAATAAAAAAGAGGCGTGGLGSQGTSGRGTLG---GQGAGAAAAA 69
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      32 MGASGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91

QY      70 AAAAAAGAGGCGGTGCTLSQGTSGRGTLGCGAGAAAAATAAAAG-CAGCGGTGCTLSQGT 128
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      92 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 151

QY      129 SGRGTLGCGCGCAAAAAATAAAAGAGCGGCGTGGLGSQGTSSPGGYGPPOQT---SGRGTLG 185
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      152 AGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 211

QY      186 GQGAAGAAAAATAAAAGAGAGGCGGTGCTLSQGTSGRG---GLGCGAGAAAAATAAAAGAG 242
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      212 GSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 271

QY      243 QGGTGTGCTLSQGTSSRGTLGCGAGAAAAATAAAAGAGCGGCGY-----GLTSQGTSGR 295
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      272 SGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 331

QY      296 GTLGGCGAGAAAAATAAAAGAGAGGCGGTGCTLSQGTSSPGGYGPPOQT---SGRGTLGCGG 352
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      332 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391

QY      353 AGAAAAATAAAAGAGAGCGGCGGTGCTLSQGTSSRG---GLGCGAGAAAAATAAAAGAGAGCGG 409
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      392 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 451

QY      410 YGGLGSQGTSSRGTLGCGAGAAAAATAAAAGAGGCGGY-----GLGSQGTSGRGTL 462
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      452 GAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 511

QY      463 GCGGAGAAAAATAAAAGAGAGCGGCGGTGCTLSQGTSSPGGYGPPOQT---SGRGTLGCGAGAG 519
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      512 AGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 571

QY      520 AAAAAATAAAAGAGAGGCGGTGCTLSQGTSSRG---GLGCGAGAAAAATAAAAGAGAGCGGCGG 576
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      572 GSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631

QY      577 LGSQGTSSRGTLGCGAGAAAAATAAAAGAGAGCGGCGY-----GLGSQGTSGRGTLGCG 629
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      632 SGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 691

QY      630 GAGAAAAATAAAAGAGAGCGGCGGTGCTLSQGTSSPGGYGP 667
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      692 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729

RESULT 11
US-08-475-411A-31
Sequence 31, Application US/08475411A
Patent No. 6140072
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-475-411A-31

```

```

Query Match      35.3%; Score 1233.5; DB 4; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3,5e-84;
Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps 11;

```

```

QY 13 MASGRGLGGGAGAAAAAAGAGGCGGCGTSGRGGLG---GCGAGAAAA 69
DB 32 MGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 91
QY 70 AAAAAAGAGGCGGCGTSGRGGLGCGAGAAAAAAGAGAGAGGCGGCGTSG 128
DB 92 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 151
QY 129 SGRGGLGCGGAGAAAAAAGAGGCGGCGTSGRGGLG---SGRGGLG 185
DB 152 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 211
QY 186 GCGAGAAAAAAGAGAGGCGGCGTSGRGGLG---GCGAGAAAAAAGAGAG 242
DB 212 GCGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 271
QY 243 GCGGCGTSGRGGLGCGAGAAAAAAGAGAGAGGCGGCGTSGRGGLG 295
DB 272 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 331
QY 296 GCGGCGGAGAAAAAAGAGGCGGCGTSGRGGLG---SGRGGLGCGG 352
DB 332 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 391
QY 353 AGAAAAAAGAGAGGCGGCGTSGRGGLG---GCGAGAAAAAAGAGAGGCGG 409
DB 392 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 451
QY 410 YGGLGSGGTSGRGGLGCGAGAAAAAAGAGAGGCGGCGTSGRGGLG 462
DB 452 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 511

```

```

QY 463 GCGAGAAAAAAGAGAGGCGGCGTSGRGGLG---SGRGGLGCGGAG 519
DB 512 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 571
QY 520 AAAAAAAGAGAGGCGGCGTSGRGGLG---GCGAGAAAAAAGAGAGGCGGCG 576
DB 572 GCGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 631
QY 577 LGSQGTSGRGGLGCGAGAAAAAAGAGAGGCGGCGTSGRGGLGCG 629
DB 632 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 691
QY 630 GAGAAAAAAGAGAGGCGGCGTSGRGGLG---GCGAGAAAAAAGAGAGGCGG 667
DB 692 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 729

```

```

RESULT 12
US-08-478-029A-31
; Sequence 31, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: Functional Recombinantly Prepared
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-029A-31

```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,509B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1985
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-7/RPT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 1101 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-509B-83

```

```

Query Match      33.7%; Score 1176.5; DB 1; Length 1101;
Best Local Similarity 40.8%; Pred. No. 5.3e-80;
Matches 270; Conservative 60; Mismatches 319; Indels 13; Gaps 7;

QY 11 GSMASGRGLG-GGAGAGAAAAAGAGGAGGCGYGLGSGGTSGRGLGGGAGAAAA 69
DB 432 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 491
QY 70 AAAAAAGAGGCGYGLGSGGTSGRGLGGGAGAAAAAGAGGCGYGLGSGGTSG 129
DB 492 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 551
QY 130 GRGGLGGGAGAAAAAGAGGAGGCGYGLGSGGTSGPG-GYGPQQTSGRGLGGG 188
DB 552 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 608
QY 189 AGAAAAAGAGAGGAGGCGYGLGSGGTSGRGLGGGAGAAAAAGAGGCGYGG 248
DB 609 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 668
QY 249 LGSQGTSGRGLGGGAGAAAAAGAGGCGYGLGSGGTSGRGLGGGAGAAAA 308
DB 669 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 728
QY 309 AAAAAAGAGGCGYGLGSGGTSGPG-GYGPQQTSGRGLGGGAGAAAAAGAG 367
DB 729 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 785
QY 368 GCGGCGYGLGSGGTSGRGLGGGAGAAAAAGAGGCGYGLGSGGTSGRGLGGG 427
DB 786 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 845
QY 428 GAGAAAAAGAGAGGCGYGLGSGGTSGRGLGGGAGAAAAAGAGGCGYGG 487
DB 846 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 905
QY 488 GLGSGGTSGPG-GYGPQQTSGRGLGGGAGAAAAAGAGGCGYGLGSGGTSG 546
DB 906 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGSG 962

```

```

QY 547 RGLGGGAGAAAAAGAGGAGGCGYGLGSGGTSGRGLGGGAGAAAAAGAG 606
DB 963 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 1022
QY 607 AGGCGYGLGSGGTSGRGLGGGAGAAAAAGAGGCGYGLGSGGTSGRGLGG 666
DB 1023 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 1082
QY 667 GQ 668
DB 1083 GR 1084

```

Search completed: July 3, 2001, 14:56:34
Job time: 524 sec

This Page Blank (uspto)

Db 371 AG---AVAAAAAGGAGGGYGGLGSQG-----AGRGQAG-----AAAAAAGAGGRYGG 419
 Qy 451 LGSQGTSGRGLGCGAGAAAAAAAAGAGCGGTYGGLGSQGT-----SGP 497
 Db 420 LGNGG-AARGGLGQGA---AAAAAAAGAGGGYGGLGNMGAGRGCGGAAAAAGAGGQ 475
 Qy 498 GGYGP-GGQTSGRRGLGCGAGAAAAAAA-----AGAGCGGTYGGLGSQGTSGRGL 550
 Db 476 GGYGGLGQGA---GGCGAGAAAAVAGQESIRQQAGCGGTYGGLGSQ--SGRGL 531
 Qy 551 GGQGAGAAA-----AAAAAGAGCGGTYGGLGSQGTSGRGLG 590
 Db 532 GGCGAGAAAAAAGAGCGGGLGCGAGCGGAGAAAAAAGVRGGYGGLGSQ----AGRGG 587
 Qy 591 OGACAAAAAAGAGAGCGGTYGGLGSQGTSGRGLGCGAGAAAAAAGAGCGGTYG 650
 Db 588 QGAG---AAAAAAGAGAGCGGTYGGLGCGV-GRGLGQAG-----AAAAAGAGCGGTYG 636
 Qy 651 GGLGS 655
 Db 637 GGVS 641

RESULT 2
 T31328
 fibroin - Chinese oak silkworm
 C:Species: Antheraea pernyi (Chinese oak silkworm)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T31328
 R:Sezutsu, H.; Tamura, T.; Yukuhiro, K.
 A:Submitted to the EMBL data library, August 1998
 A:Description: Characterization of the full length fibroin gene of a wild silkworm, Ant
 A:Reference number: Z20995
 A:Accession: T31328
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-2639 <SEZ>
 A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
 A:Genetics:
 A:Introns: 14/3

	Query Match	39.7%	Score 1386.5;	DB 2:	Length 2639;
Db	Best Local Similarity	36.7%;	Pred. No. 9e+71;		
	Matches 414; Conservative	25;	Mismatches 210;	Indels 479;	Gaps 39;
Qy	14 ASGRGLGCGQ-----AGAAAAAAAAAGAGCGGTYGGLGSQGTSGRGLGCGAGAAA 67				
Db	468 ASGAGSGGTYGSDSAAAAAAAAAACAGAGAGC--GSYGM-GDGYGSDSAAAAA 524				
Qy	68 AAAAAA---AGAGCGGTY----- 83				
Db	525 AAAAAAASGAGRGDDGSGSSAAAAAAAAAARARRHDSNAASAAAAA 584				
Qy	84 --GLGSQTS--GRGLGCGAGAAAAAAGAGCGGTY----- 122				
Db	585 ASGAGSGGTYGWDGTYSDSAAAAAASGAGAGGTYGWDGTYGSDSAAA 644				
Qy	123 -----LGSTSGRGLGCGAGAAAAAAGAGCGGTY-----GGLSQ 164				
Db	645 AAAAAAASGARSGGTYGSDSAAAAAAGAGAGGTYGWDGTYGSD 704				
Qy	165 -----GTPGGTYGPQ-----GTSGRGLGCGAGAAA 193				
Db	705 SAAAAAASGAGRGDGTYGSSAAAAAASAARRAHDSNAASAAAAA 764				
Qy	194 AAAAAAGAGC--QGYGGLGSQGTSGRGLGCGAGAAAAAAGAGCGGTYG--- 248				
Db	765 AAAAAAASGAGSGGTYG---GDGYGSDSAAAAAAGAGAGGTYG 817				
Qy	249 -----LGSTSGRGLGCGAGAAAAAAGAGCGGTYG 285				
Db	818 WGDGYSDSAASAAAAAASGARSGGTYGSDSAAAAAAGSDAGVG 877				

[illegible]

RESULT 3
 F70806
 hypothetical glycine-rich protein RV3508 - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: F70806
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajendram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70806
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1901 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:93261554; PIDN:CAA17745.1; PID:92922
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: RV3508
 C:Superfamily: collagen alpha 1(IV) chain

[illegible]

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70846
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno-
A:Reference number: A70500; MUID:98295987
A:Accession: B70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1538 <COL>
A:Cross-references: GB:AL021841; GB:AL123456; NID:93261517; PIDN:CAA17117.1; PID:g326
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3345C
C:Superfamily: collagen alpha 1(IV) chain

	Query Match	Similarity	33.1%	Score 1157:	DB 2:	Length 1538:	
	Best Local	Similarity	34.4%	Pred.	No. 4e-58:	Mismatches 308:	Indels 260; Gaps
	Matches	317;	Conservative	36;			
Qy	1	MKGSHHNHHSMSAGCGGLGCGAGAAAAAAG-----AGCGYGLCSOGTS	53				
Dd	549	LNGSGANGPTPTSGKNGNCGAGATTTVAEGEAGCAGNGCHGGSVGNCGAGAGANGTA	608				
Qy	54	G----RGLGCGGA-GAAAAAAAAGAGCGYGGLSQ-----GTS	91				
Dd	609	GTGALLNGNGNGGIGGNGSAAAGTGDDGCKGAGAGANQDFSAANGANCGGNG	668				
Qy	92	GRGGLGCGAGAAAAAAGCAGCGYGLCSOGTSRGGLGCGGA-----	139				
Dd	669	GNGCTGCGGDAFTFAFKAGNGAGGNG-----GNVGAAGCGAGKGALPAMKATGADG	724				
Qy	140	-----CAAAAAAAGAGCGYGG--LGSGTSGPGCYPRGO-----T	178				
Dd	725	TAPTSGDGGNGNCASPTVAVGNGGCGGSSGVNGNGNCAGACGNAAQACTPCPT	784				
Qy	179	SGRGGLGCGAGAAAAAAGS--AGCGYGLCSOGTSRGGLGCGAGAAAAA	236				
Dd	785	SDDSSTSCSTDGCAGAGNGCAGGAGGTLAGHGNCGKK--GNGCGGTLGGAENGAGAGCN	842				
Qy	237	AAGGAGCGCGYGLCSOGTSRGRLGSG--GAGAAAAAAGAGAGCGYGG-----	286				
Dd	843	ANGANGENG-----GSGNGGCGGAGAGNGAKRAQAAYTDATGTGDBGNGCDGKAG	898				
Qy	287	-----LGSOGTSGRGLGQQ-----	301				
Dd	899	DGAGENGLNSGAMLPGGCTVGNPCTGANGNGCNAVGCTGCKACTGSLTGLDTGTT	958				
Qy	302	-----GAGAAAAAAG-----AGCGYGG-----LSOGTS	331				
Dd	959	PNGGNGNGNGCKGKTGANGSGAAGNGGSGGLNGPDAGNGGAGALNQAFPGTIG	1016				
Qy	332	GPBGY-----GPGQQTSGRGGLGSGGAGAAAAAAGAGCGYGLCSOGTSRGGLG	387				
Dd	1019	GGGNGNGMGAMINGLJGFGGAGGGGADVAAITTGAGAGNGGAGGPASTGLGEPGAG	1077				
Qy	388	GCGAAGAAAAAAGGA-----CGCGYGLCSQS-----TSRGGLGCGGAGAAAAA	436				
Dd	1079	GRGAGADRPSGYGAGAGCGAGAGYGTGCGCGGICGBERTGANGSGSDGGGGISLGG	1133				
Qy	437	AAAAG-----AGGCGYGLCSOGTSRGGLGSGGAGAAAAAAGAGAGCG	484				
Dd	1139	NGLGNGNGVSFETRCGAGANGYRGPG--GPENGGLGNGGAGNGGISTGCGDGAG	1196				
Qy	485	GYGGLGSOGTSPPGYGPGQQTSGRGGLGCGAGAAAAAAG-----AG	532				
Dd	1197	GGGNGGD--GNVLGJDADSGAGGNGGIGTIDACAGAGAGAGNGGSSKSTTTGNAG	1255				
Qy	533	CGGYGLGSGTSGKGLGCGGAGAAAAA-----AAGAGAGCGYGLCSOGTS--GRGL	588				

[illegible]

```

Db      556  GGNMGAKGTFDNGGDCGVNGNGSGRCITGAGGIGTGAAGTACADGACRATPTTSGGNGGTG  615
Oy      354  --GAAAAAAGAGAGAGCGYGL-GSGTSGRGLGCGAGAAAAAAGAGAGG---  407
Db      616  GNGCAATYAGGAGGAGCGGNGGLVGNCGAGGKGGDGAAGVAGSSPTTAGBSGTSGGNGG  675
Oy      408  -GTYGGLCSQ-----GTSGRGLGCGGAGAAAAAAGC-----AGCGY  448
Db      676  AGGAGAGAGRGDFGCGDGTGGAGGNGANGNATTPGAKGGDGGHGPAGAGCGNGCGGP  735
Oy      449  GGL-----GSGTSGRGLGCGGAGAAAAAAGAGAGCGGYG---GLGS  491
Db      736  GGLAENLFQGNGTQGVGSSGCKGAGAGLAGDGNKANPNFAPGDNNGHGNGNPNPAGG  795
Oy      492  QGTSGPGYGPQQ-----TSGRGLGCGGAGAAAAAAGAGAGCGYGL-GSQ  542
Db      796  QGSGGAGSTPGAKGAFPTPTSGDGGDGGNGN-----SQVVGNGCGDGGNGNGSSA  850
Oy      543  GTSRGLGCGGAGAAAAA-----AAAGAGCGGYGLGSLGSGTSGNGG  587
Db      851  GTGGNGGRCGDAFEGMSANATNPGENPNPNGNGAGAGAGGLNG--GNGAGGNGG  909
Oy      588  LGGCGAGAAAAAAGAGAGG---GGYGLG-----SGTSGRGLGCG  629
Db      910  LGGFGGNGAGAGYAVAPGQPGSAGGHGAGNGAGCGGCVSDAGAGAGAGCD  969
Oy      630  G-----AGAAAAAAGAGAGCGGYG-LGSQGTSGPGYGPQQTSGIRPPAKL  680
Db      970  GGAPDGNANGCGAGAFAGGAGGCGGDDGCGNAGNAGCGPGGTG---STGKAGPAGSI  1026

RESULT  11
A:0896
A:0896
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: A70896
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
  ; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
  Rajadaram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
  Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
  A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
  A:Reference number: A70500; MUID:98295987
A:Accession: A70896
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <COI>
A:Cross-references: GB:A7021897; GB:AL123456; NID:q3256022; PIDN:CAA17207.1; PID:el259
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv1091
C:Superfamily: unassigned collagens

Query Match      32.6%; Score 1139; DB 2; Length 853;
Best Local Similarity 40.3%; Pred. No. 2.6e-57;
Matches 298; Conservative 27; Mismatches 304; Indels 110; Gaps 28;

Oy      11  GSMASGRGLGCGAGAAAA-----AAAAAGAGCGGYGLGSGTGS-----GRGGL  58
Db      137  GGLLVGNGNGSTTTAGVAGNGAGLIGNGAGGGGAGAAAGNCGAGGWLTYGNNGA  196
Oy      59  GGCGGAAAAAAGAGAGAG-----GGGYGGLGSGTSG-----RGGLGCGAGAAAA  107
Db      197  GGAGGTSVIPGVAGNGAGAGSAGLMTGTGAGAGDGNRSGSVNVVAGSAGNGAGGAAG  256
Oy      108  AAAAAGAGGCGGYGLG-----SGTSGRGLGCGGAGAAAAAAGAGAGCGGYG  159
Db      257  LFGDAGAGNGGKAGAGAPSFINTAGDAGAGAGGSGGHAHLMLGAGAGAGNGSGGTG  316
Oy      160  GIG--SGTSGPGYGPQQGTS---GRGGLGCGGA---GAAAAAAGAGAGCGGYG  209

```


[illegible]

A:Cross-references: GB:Z59890; GB:AL123456; MID:g3242245; PID:CAB09322.1; PID:g213131
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV1759c; wag22
C:Superfamily: elastin

```
Query Match          31.7%; Score 1106.5; DB 2; Length 914;  
Best Local Similarity 38.2%; Pred. No. 1.8e-55;  
Matches 306; Conservative 31; Mismatches 304; Indels 159; Gaps 32;
```

OY 11 GSMAAGRGGLGGGAGCAAAAAAAAAAAAGAG----QGYYGGL-GSGGTSGRGLGGGAGA 65
DB 116 GNGANGAPGTGNMGDDACMLIGNGAGSSGAKGANGAGPEGAAGLFNGAGAGSTA 175
OY 66 AAAAAAAAAGGA-----GQGYGGLGSQTGS--GRGGLGGGAGAAAAAAAAGA 115
DB 176 TANNIGTAGAGAGSNAMLEFGAGGAGAGGAATSLVGIQGTCTGNNMGLAANAAGA 235
OY 116 G-----QGYYGGLGSQ-----TSGRGGLGGG-----AGAAAAAAAAGG-AGQGYG 159
DB 236 GFPSFTAGAGAGAGAGAGLFTTGVGVGAGCGHTTGAGAGAGAGAGLFCAGCMGAGCFG 295
OY 160 GLGSGGTSPGTYGTCQQTSGRGLGGGAGAAAAAAAAGC-----AGQGYG 209
DB 296 DHGTGTGTGAGADGGGGGLFGAGSDGAGSGSLTTGCAAGNAGNAGTLSLGANAGAGGTG 355
OY 210 GLGSGT---SGRGGLGGGAGAAAAAAAAGAGAG-----QGYYGGLGS-----QGTSG 256
DB 356 GAG--GYTFGGKKGKAGAGAGNAGMLFGSGGGGTGCGTFAAGCGGTYGGSAGMLSGSG 413
OY 257 RGLGGGQ--AGAAAAAAAAGAGAGGAGGGLGGSQ--TSGRGGLGGGAGAG-----A 306
DB 414 SGGAGGSGGPACTAAGAGAGAGAGPGLIGNGNNGNGSGSGTGVGAGNAVLIGNG 473
OY 307 AAAAAAAGCAGCGYGGL-----325
DB 474 EGGIAGLAGSKSFGRFGGILLGADYNAPESTSPWHNLQDILSFINEPTALTGRPLIG 533
OY 326 -GSQGTSGPG--GYGQQQTSGRGLGGGAGAAAAAAAAGAG-----GQGYGGLSQ 378
DB 534 NCDSTPGTGDGAGAGMLFENGNGAGAGA-AGYNGSAGAGAGAGLLFTGTGAGAGAGCV 592
OY 379 GTSGRGLGGGAGAAAAAAAAGAGAGGAGGAGGGLGSGTSGRGLGGGAGAAAAAAA 438
DB 593 GTAAGAGAGAGAGSNFL-----ISGCTGCVGAATT-TGTVGAGGNAAGLLIGA 642
OY 439 AAGAGAGGAY--GELDSQGTSGRGLG-----QGAAAAAAAAGAGAGGAGGGLGSG 492
DB 643 GLGGCGCGCAFATTGVTTCGAGTGTGAAGLFPANAGACAGAGCTGSTAGAGAGAGGLTAHG 702
OY 493 GTSGGCGYPRGQOTSGRGLGGGAGAAAAAAAAGAGAG--QGYYGGLGSQGT----- 545
DB 703 GTGGPRGNNGSTGAGCTGGAGPG--GLYGAGSSGAGGGHAGAGGGGVGAGNAGSLTTANA 761
OY 546 --GREGLAGQ-----GAGAAAAAAAAGAGAGGAGGAGGLGSGTSGRGLGGG----- 592
DB 762 SGAGAGSGSSLSLKRAGAGAGAGSAGLFTYGSAGAGAGNGYSLSNGTGGDBGTTGCAQITGL 821
OY 593 ---AGAAAAAAAAGAGAGGAGGAGG-LGSGTSGRGLGGGAGAAAAAAAAGAGAG 647
DB 822 RSFGGAGAGAGASPTGAAGNGAGAGKAGLYNGGDDGAGAGDGATSGKAGAGNAVIVIGN 881
OY 648 GGYGGLGSQ--GTSRPGCYG 665
DB 882 GGNGAGNAGAGGTAGAGAG 901

RESULT 15
B70812
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text change 20-Jun-2000
hypothetical glycine-rich protein rv0834c - Mycobacterium tuberculosis (strain H37RV)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:32 ; Search time 37.35 Seconds

(Without alignments)

624,578 Million cell updates/sec

Title: US-09-490-291-4

Sequence: 1 MRGSHHHHHHSGMSASRGGL.....GCGPQGQOTSGIRPAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2135	61.1	747	1 SPOL_NEPCL	P19837 nephila cla
2	1277.5	36.6	1901	1 Y208_MYCTU	O53533 mycobacteri
3	1158	33.4	5263	1 FBOH_BOMO	P05780 bombyx mori
4	1135	32.5	627	1 SPD2_NEPCL	P46804 nephila cla
5	1106.5	31.7	914	1 WA22_MYCTU	O06794 mycobacteri
6	1044.5	29.9	778	1 YQ34_MYCTU	P71933 mycobacteri
7	1028	29.4	957	1 Y278_MYCTU	P56877 mycobacteri
8	988	28.3	801	1 Y747_MYCTU	O53810 mycobacteri
9	976	27.9	860	1 ELS_MOUSE	P54320 mus musculu
10	939	26.9	864	1 ELS_RAT	O99372 rattus norv
11	869	24.9	603	1 YD25_MYCTU	O10637 mycobacteri
12	853.5	24.4	465	1 GRP2_PHAVU	P10496 phaseolus v
13	775.5	22.2	747	1 ELS_BOVIN	P04985 bos taurus
14	763	21.8	641	1 EBN1_EBV	P03211 epstein-bar
15	704.5	20.2	750	1 ELS_CHICK	P07916 gallus gall
16	691	19.8	730	1 ELS_HUMAN	P15502 homo sapien
17	686	19.6	498	1 Y118_MYCTU	O50615 mycobacteri
18	679	19.4	1733	1 VNDA_PPKA	P33485 pseudobacile
19	678	19.4	543	1 YP91_MYCTU	O50603 mycobacteri
20	673	19.3	384	1 GRP1_PETHY	P09789 petunia hyb
21	669	19.2	481	1 LORT_MOUSE	P18185 mus musculu
22	665	18.8	1027	1 CAEP_RIFPA	P30754 riftia pach
23	644	18.4	338	1 GRP_ARATH	P27483 arabidopsis
24	636	17.9	515	1 Y140_MYCTU	O50594 mycobacteri
25	625	17.9	1373	1 CA21_MOUSE	O01149 mus musculu
26	621	17.8	1466	1 CA13_HUMAN	P02461 homo sapien
27	616	17.6	434	1 YK96_MYCTU	O10707 mycobacteri
28	608	17.4	1453	1 CA11_CHICK	P02457 gallus gall
29	607.5	17.4	1418	1 CA12_HUMAN	P02458 homo sapien
30	604.5	17.3	1460	1 CA11_CANFA	O95817 bos taurus
31	603.5	17.3	1409	1 CA13_BOVIN	P04258 bos taurus
32	602.5	17.2	463	1 YA68_MYCTU	O53416 mycobacteri
33	602	17.2	1464	1 CA11_HUMAN	P02452 homo sapien

34	598	17.1	1459	1 CA12_MOUSE	P28481 mus musculu
35	597.5	17.1	1464	1 CA13_MOUSE	P08121 mus musculu
36	597	17.1	1366	1 CA21_CANFA	O46392 canis famli
37	596	17.1	1366	1 CA21_HUMAN	P08123 homo sapien
38	595	17.0	1363	1 CA21_BOVIN	P02465 bos taurus
39	594.5	17.0	1453	1 CA11_MOUSE	P11087 mus musculu
40	573.5	16.4	1362	1 CA21_CHICK	P02467 gallus gall
41	563	16.1	1758	1 CA24_CAEEL	P17140 caenorhabd
42	568	16.0	1367	1 ICEN_XANCT	P18127 xanthomonas
43	557.5	16.0	1758	1 CA14_MOUSE	P17139 caenorhabd
44	555	15.9	1775	1 CA14_DROME	P08120 drosophila
45	554.5	15.9	779	1 CA11_BOVIN	P02453 bos taurus

ALIGNMENTS

```

RESULT 1
SPOL_NEPCL STANDARD; PRT: 747 AA.
ID SPOL_NEPCL
AC P19837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxId=6915;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 653-747 FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
  proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
  Araneus bidentatus (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
CC - FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE
  CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
  PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
  WITH ELASTIC AMORPHOUS SEGMENTS.
CC - SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC - SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37137; AAA29380.1; -
DR EMBL: U03848; AAB60212.1; -
DR PIR: A36068; A36068.
KW Silk; Repeat.
FT DOMAIN 1 1 655 25 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 1 1 25
FT REPEAT 26 38
FT REPEAT 39 66
FT REPEAT 67 96
FT REPEAT 97 130
FT REPEAT 131 158
FT REPEAT 159 191
FT REPEAT 192 204
FT REPEAT 205 235

```


Db 412 GTGAGAGAGGCGGCGAATAGTNGSGAGGTGGGAGGCGMGGSGADNAGSGAGG 471
 QY 137 OG--AGAAAAAAGG-----AGGGTGGGAGGCT----- 166
 Db 472 TGGNAGGAGGAGAAAGTGTGTVGAAGKACIGGTGGGAGGAGSAGTADATGATGCTG 531
 QY 167 -----SGPGYGPQGTSGRGLGCGGAGAAAAAAGAGAG-----GGYGLGSG 215
 Db 532 FSGAGAGAGAGAGTGGTGGTNGSGGGG--GTGAGAGAGAGAGGADNPTGTGGTGTGGK 590
 QY 216 TSGRGLGCGG--AGAAAAAAGG--AGGGYGLGSGT--SGRGLGCGAGAGAA 269
 Db 591 --GAGAGCGGGGAGAGCTGSGAGGTGGGAGAGAGAGADPTGTGAGGTGTGGA 648
 QY 270 AAAAAAGAGGCGGAGG-----LGGGTSGRGLGCG--GAGAAAAAAGAG-- 317
 Db 649 AGAGAGAGAGAGTGTGAGVSVNAGTGTGTGGTGGAGAGAGAAAGSAGTGGAGFAG 708
 QY 318 GGGGAGGAGG-----GSGPGYGPQGTSGRGLGCGA-----GAAAAA 360
 Db 709 GAGGAGAGAGAGTGGTGGTNGSGAGAGAGCGGTGAGAGSADNPTGTGAGAGGTGGAAG 768
 QY 361 AAGAGAG--GGGAGGGL-----GSGGTSGRGLGCGA----- 391
 Db 769 AGAGAGAGTGTGTGGTGGTNGSGAGAGAGCGGTGAGAGSADNPTGTGAGAGGTGGAAG 828
 QY 392 GAAAAAAGAGAG--GGGAGGAGGAGTGG--RGLGAG--GAGAG-- 433
 Db 829 GSAGAGAGTGGAGAGAGAGCGGAGGAGTGAAGLGDNGAGVGGAGAGAGAGAGAGAGVGLTAK 888
 QY 434 AAAAAAGAGAGGAGGAGGAGGAGG-----TSGRGLGCGA-----GAAAAAAGAGAG-- 481
 Db 889 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGTGTGAGAGAGAGAG 948
 QY 482 GGGGAGGAGGAGTGTGAG 541
 Db 949 GTGGAGAGAGAGTGTGAG 1005
 QY 542 GGTG-----GAGGAGGAG 587
 Db 1006 DGASGAG 1065
 QY 588 LGAG--GAGAG--AAAAAAG 629
 Db 1066 VGGAG 1125
 QY 630 GA-----GAAAAAAG 665
 Db 1126 GGGGAGGAGAGTGTGTGAG 1169

RP SEQUENCE OF 1-168 FROM N.A.
 RX MEDLINE=80045039; PubMed=498286;
 RA Tsujimoto Y., Suzuki Y.;
 RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
 RT flanking, mRNA coding, entire intervening and fibroin protein coding
 RT regions.";
 RL Cell 18:591-600(1979).
 RN [13]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=79211211; PubMed=455439;
 RA Tsujimoto Y., Suzuki Y.;
 RT "Structural analysis of the fibroin gene at the 5' end and its
 RT surrounding regions.";
 RL Cell 16:425-436(1979).
 RN [14]
 RP PARTIAL SEQUENCE FROM N.A.
 RX STRAIN=KINSHU X SHOMA;
 RX MEDLINE=89094868; PubMed=3210244;
 RA Mita K., Ichimura S., Zama M., James T.C.;
 RT "Specific codon usage pattern and its implications on the secondary
 RT structure of silk fibroin mRNA.";
 RL J. Mol. Biol. 203:917-925(1988).
 RN [15]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=94355842; PubMed=7916056;
 RA Mita K., Ichimura S., James T.C.;
 RT "Highly repetitive structure and its organization of the silk fibroin
 RT gene.";
 RL J. Mol. Evol. 38:583-592(1994).
 RN [16]
 RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
 RX STRAIN=J-139;
 RX MEDLINE=99296390; PubMed=10366732;
 RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
 RA Takagi T., Mizuno S.;
 RT "Determination of the site of disulfide linkage between heavy and
 RT light chains of silk fibroin produced by Bombyx mori.";
 RL Biochim. Biophys. Acta 1432:92-103(1999).
 CC -1- FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
 CC -1- SOLUBLE AND CHEMICALLY INERT FIBRE.
 CC -1- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
 CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
 CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
 CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSS)
 CC SECTION OF SILK GLANDS.
 CC -1- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
 CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
 CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
 CC ALA-GLY-ALA-)N INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
 CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
 CC AMORPHOUS REGIONS.
 CC
 CC *****
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC *****
 CC EMBL: AF226688; AAF76983.1; -
 CC EMBL: V00094; CA23432.1; -
 CC EMBL: V00097; CA23433.1; -
 CC EMBL: S74439; AAB31861.1; -
 CC EMBL: X13869; CA23076.1; -
 CC EMBL: M35378; AA27839.1; -
 CC EMBL: AB017362; BAA33147.1; -
 CC PTR: S01844; S01844.
 CC SILK: Signal; Repeat.
 KW SIGNAL 1 21
 FT CHAIN 22 5263 POTENTIAL.
 FT DOMAIN 149 5206 FIBROIN HEAVY CHAIN.
 FT DISULFID 5244 5244 HIGHLY REPETITIVE.
 FT INTERCHAIN (WITH LIGHT CHAIN).

D**b** 495 SCGAGAVTGGNGAGCGNMGALLGAAAGACGAGAGAVG--GNGCGAGNG-GLFANGGAGCGP 458

Q**y** 404 G----AAQGGTGGGGSQ-CTSRGRCGCGCGCAAAAAAAAAAAGGAGGGGTGGGSGQTSG 458

D**b** 552 GFGSPAAAGTGGGAGGNGGNGFAGAGCTGAGAGGSLTLACGAGAGGNG-----GLFG 601

Q**y** 459 RGGTGGCGAGAAAAAAAAAAGAGAGGGGGLGSGQTSFGPGYGPQDTSGRGGLGCGAG 518

D**b** 602 AGCTGAGSHSTAAVAGCGAGGA--CGDACLTLTGAGSGGAGSGGSSLTAAGVVGTG-- 657

Q**y** 519 AAAAAAAAAAGACGCGGTGGGSGQTSGSRGGLGSGQ-----GAGAAAAAAAAAAGAGACGG 573

D**b** 658 ---GAGGTLFGSGGAGSGGFSNNGNGAGAGGAGGDAAGLVGSGAGGAGASATGAATTGG 713

Q**y** 574 YGGGSGQTSGRGGLGSGGAGAAAAAAAAAAGAGAGGGTGGGGLGSGQTSGRGGLGCGAGA 633

D**b** 714 DGGAG--GRSGAFGLGGDGGCGAGTGLSGAFHIGKGGVVG--GSAVLIGNGNGNGNGNS 768

Q**y** 634 A-AAAAAAAAAGAGGGGCGG--LTSQTS 660

D**b** 770 GNAGSGGAPGPGAGGAGGILLLENGING 799

RESULT	9
ELN_MOUSE	
AC	ELN_MOUSE
ID	STANDARD;
AD	PRT; 860 AA.
CD	P54320;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	ELASTIN PRECURSOR (TROPOLASTIN).
GN	ELN.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_Taxid:10090;
RA	[1]
RA	SEQUENCE FROM N.A.
RP	STRAIN=BALB/C; TISSUE=Lung;
RX	MEDLINE=95130069; PubMed=7829060;
RA	Wynner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT	"Use of an intron polymorphism to localize the tropoelastin gene to
RT	mouse chromosome 5 in a region of linkage conservation with human
RT	chromosome 7.";
RL	Genomics 23:125-131(1994).
CC	-1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC	NOCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC	-1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC	INTO AN EXTENSIBLE 3D NETWORK.
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC	-1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the swiss institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC	or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL; U08210; AAA80155.1; -.
DR	HSSP; P04002; IWEA.
DR	MGD; MGI:95317; Eln.
KW	Structural protein; Repeat; Signal; Connective tissue.
FT	SIGNAL 1
FT	CHAIN 28 860
FT	POTENTIAL.
FT	ELASTIN.
SO	SEQUENCE 860 AA; 71955 MW; 0CB055AAE1EDD7F1 CRC64;

Query Match	27.98;	Score 976;	DB 1;	Length 860;
Best Local Similarity	39.38;	Pred. No. 3.4e-37;		
Matches 323;	Conservative 36;	Mismatches 249;	Indels 214;	Gaps 40.

OY	11	GSMAASGKGLGGGG-----AGAAAAAAAAAGAGGGG	46
Db	78	GTFGAGGCGGLGAGCPGAGLCAFPAGTFFPGAGALVPGGMAAGMAAAYAAAKAGAGLGGVGG	137
OY	47	LGSGGTSGRGL- <i>CGCGAG</i> AAAAAAAAAAAGCAGCGGTGGLGSGGTSGRGLGGCGAGAA	105
Db	138	V--PGGVGSGVPGGVG-----GVPGGVGGVPGP-----GVGGIGGIGLGVSTGA	184
OY	106	AAAAAAG-GAG--QGCTGGLGSGGTSGRGLGGCGA-----GA	149
Db	185	VVPVGGAGIAGGCGKPGKVPVGLPGVPGGLVLTGARPPGVCLPGLPTGCVAKAPG	244
OY	150	AGAGGQ--GYGGLGSG-----GTSPPGGY-----PGQTSGRGLGGQ-----	187
Db	245	GGGAFSGIPGVPGPGGQQPVPVPLCTPKAKKLPDGTGVLPTYNKKLTYVAAGAGKAGYPT	304
OY	188	--GAGAAAAAAAAAGAGGCGYGC--GLGSGTSGR-----GGLCGCGAGAAAAA	235
Db	305	GTVGSGQAAAAAAAKAAYGAGAGVLPFGVGGGILPGGAGAPIGIGTIGAGCTPPAAAAAAK	364
OY	236	AAAGCA-----GQGGYGGIAGSGGTSGRGLGGGOGAG	266
Db	365	AAAAAAKYGAAGGLVPGGPGVRLPGAGIPGVGGLPGVGGILPGVGGGILGSPGILVGGPAAV	424
OY	267	AAAAAAAAAAGCA---GQGGYGC---GLGSGTSGRGLGGGOGAGAAAAAAGAGAG	319
Db	425	SPAAAAKAAAKAAAYGARGCGIGITTYGVGAGGPGYGVGAGAGLGGASPAAAAAAAKAK	484
OY	320	GGYGGLSQGTSGG---GYGPG--QOTSRRGLGGCGA	368
Db	485	YGAGAGAGLGLVGAVALPGALPGVAPVPAGGVPAGCTPAAAAA	544
OY	369	QGGY-GGLGSGGTSGRGLGGGCGA	427
Db	545	VGGVPGGTVGCTGGVGVG-----VPGGVGGPVGTGIC---A	588
OY	428	GAGAAAAA-----AAAG-GAGCGYGC-GLGSGTSGRGLGGCGAGAAAAA	476
Db	589	GSPAAKAAKAAKAAKAYRAAAGLGAAGVPGGAGAGVPGGAGAGVPGFAGAGVPGFGA	648
OY	477	AAG-----GAG-----GGYGGIAGSG--GTSPPGGYGGQOTSRRGLGGG--	516
Db	649	GAGVPGFAGAVPGSLAASKAAKGAAGGGLGPRGGLGPRGLG-----GPGILGAGVP	702
OY	517	---AGAA-AAAAAAGAGAGCGGYGGLGSGGTSGRGLGGGAGAAAAA	572
Db	703	GRVGAAPAPAAAAAAKAAKAAKAAQYGLGAGGLGAGLGAAGL-----AGLIGAG	753
OY	573	GY--GGLGSGTSGRGLG-GGAGAAAAAAGAGAGCGGTGGLGSGGTSGRGLGGG	629
Db	754	GIGAGGLGAGGL-GAGGIGAGAGGVSPAAAAAAAYGAAGLGGV--LGARPPGGGVAAAP	810
OY	630	GAGAAAAAAGAGAGGQ-----YGG-LGSGTSGPGYGC	665
Db	811	GFGLSPTIPGGGAGGLGVGGFPFYPYGGALGALGTGGCGCTG	852

RESULT	10		
ELS_RAT			
ID	ELS_RAT	STANDARD:	PRT; 864 AA.
AC	Q99372;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-OCT-2000 (Rel. 40, Last annotation update)		
DE	ELASTIN PRECURSOR (TROPOLASTIN) (FRAGMENT).		
GN	ELN.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91104868; Pubmed=1702999;		


```

Db 191 AG-----GGAG-GGYGGGEGHGGGGGGGGG-GAGGGYAGAGHGGAGGGGGG 238
Qy 325 LGSGCTSGPGGYPGQQTSGRGLGCGAGAAAAAAGAGAGGCGYGLGSGTSGRG 384
Db 239 AG-----GGYGAGGEGHGGAG-GGCG-----GGAG-----GGYGAGGEGHGG 274
Qy 385 GLGCGAGAAAAAAGAGAGGCGYGLGSGTSGRGLGCGAGAAAAAAGAGAG 444
Db 275 AGGCGG-----GGAG-----GGYGAGGEGHGGAGGGGG-----GGAG 306
Qy 445 QGCGGCGTSGRGLGCGAGAAAAAAGAGAGGCGYGLGSGTSGPGYGRGQ 504
Db 307 -----GGTGAAGGEGHGGGGGGGGGAGGAGYAAVGEHG-----GGYG-----GGCGGGDGGGTCGCG 357
Qy 505 QTSRGLGCGAGAAAAAAGAGAGGCGYGLGSGTSGRGLGCGAGAAAAAAGAG 564
Db 358 EHGGGCGG-GCGGGG-----AGGGYGTGEGHGGGCG-----GGGGGGGCGAG----- 398
Qy 565 AAGGAGGCGYGLGSGTSGRGLGCGAGAAAAAAGAGAGGCGYGLGSGTSGRG 624
Db 399 --GCGGAAGYGG-----GEGGGGGGSGG-----YGDGGAHGGGCG-----GAG 435
Qy 625 GLGCGAGAAAAAAGAGAGGCGYGLGSGTSGPGYCG 665
Db 436 GGGGCG-----AGGAHGGGCGGGGCG-----GGHG 461

RESULT 13
ELS_BOVIN STANDARD: PRT: 747 AA.
ID P04985; P04986; P04987; Q29421;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ELASTIN PRECURSOR (TROPOELASTIN).
OS ELN. taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID:9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87194772; PubMed-3032943;
RA Raju K., Anwar R.A.;
RT "Primary structures of bovine elastin a, b, and c deduced from the
RT sequences of cDNA clones".
RL J. Biol. Chem. 262:5755-5762(1987).
[2]
RN [2]
RP SEQUENCE OF 1-27 FROM N.A.
RC TISSUE=Nuchal ligament;
RX MEDLINE-89274159; PubMed-2543440;
RA Yeh H., Anderson N., Ornstein-Goldstein N., Bashir M.M.,
RA Rosenbloom J.C., Abrams W.R., Indik Z., Yoon K., Pairs W., Mecham R.,
RA Rosenbloom J.;
RT "Structure of the bovine elastin gene and SI nuclease analysis of
RT alternative splicing of elastin mRNA in the bovine nuchal ligament.";
RL Biochemistry 28:2365-2370(1989).
[3]
RN [3]
RP SEQUENCE OF 1-27 FROM N.A.
RX MEDLINE-91234332; PubMed-2031719;
RA Manohar A., Shi W., Anwar R.A.;
RT "Partial characterization of bovine elastin gene; comparison with the
RT gene for human elastin".
RL Biochem. Cell Biol. 69:185-192(1991).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NOCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A (SHOWN HERE), B AND C; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.

```

```

CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02717; AAA30503.1; -
DR EMBL; K03505; AAA30505.1; -
DR EMBL; K03506; AAA30506.1; -
DR EMBL; J02855; AAA30776.1; -
DR EMBL; M58652; AAA03519.1; -
DR PIR; A26728; A26728.
DR PIR; B26728; B26728.
DR PIR; C26728; C26728.
DR HSSP; P04002; IWFA.
KW Structural protein; Connective tissue; Repeat; Signal;
KW Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 1 747
FT MOD_RES 105 105 ELASTIN.
FT MOD_RES 109 109 OXIDATIVE DEAMINATION.
FT MOD_RES 252 109 OXIDATIVE DEAMINATION.
FT MOD_RES 271 271 OXIDATIVE DEAMINATION.
FT MOD_RES 275 275 OXIDATIVE DEAMINATION.
FT MOD_RES 324 324 OXIDATIVE DEAMINATION.
FT MOD_RES 327 327 OXIDATIVE DEAMINATION.
FT MOD_RES 400 400 OXIDATIVE DEAMINATION.
FT MOD_RES 404 404 OXIDATIVE DEAMINATION.
FT MOD_RES 407 407 OXIDATIVE DEAMINATION.
FT MOD_RES 448 448 OXIDATIVE DEAMINATION.
FT MOD_RES 489 489 OXIDATIVE DEAMINATION.
FT MOD_RES 493 493 OXIDATIVE DEAMINATION.
FT MOD_RES 544 544 OXIDATIVE DEAMINATION.
FT MOD_RES 548 548 OXIDATIVE DEAMINATION.
FT MOD_RES 552 552 OXIDATIVE DEAMINATION.
FT MOD_RES 606 606 OXIDATIVE DEAMINATION.
FT MOD_RES 609 609 OXIDATIVE DEAMINATION.
FT MOD_RES 645 645 OXIDATIVE DEAMINATION.
FT MOD_RES 649 649 OXIDATIVE DEAMINATION.
FT MOD_RES 685 685 OXIDATIVE DEAMINATION.
FT MOD_RES 688 688 OXIDATIVE DEAMINATION.
FT MOD_RES 226 239 MISSING (IN ELASTIN B).
FT VARSPLIC 226 239 MISSING (IN ELASTIN C).
FT CONFLICT 2 3 RS -> G (IN REF. 2 AND 3).
FT CONFLICT 12 12 E -> G (IN REF. 2 AND 3).
SO SEQUENCE 747 AA; 64229 MW; 633C03E411643D83 CRC64;

Query Match 22.2%; Score 775.5; DB 1; Length 747;
Best Local Similarity 37.9%; Pred. No. 1.9e-28;
Matches 288; Conservative 26; Mismatches 282; Indels 163; Gaps 36;

Qy 18 GLGCGAGAAAAAAGAGAGGCGYGLG--SGTSGRGLGCGAGAAAAAAGAG 74
Db 27 GGVPGAVPGVPGVPGVFPFGAGLGLGVLGPGVAPRPGVGLGEGSLP-- 83
Qy 75 GAGGCGYGLGSGTSGRGLGCGAGAAAAAAGAGAGGCGYGLGSGTSGRG 134
Db 84 GAFFGGR-----FCAGGGAAGAAAYKAAAGAGAGLGVG1-----GGV 123
Qy 135 GCGAGAAAAAAGAG--GAG--GCGYGLGSGTSGPGY--GPGQOTSGRGLGCG 189
Db 124 GGLGVSTGAVVPQLGAGYAGYKPKVPGVGLPVY--FGVLPAGAGARPGVLPVPT 182
Qy 190 GA-----AAAAAAGAGAGG-----GYG-----GLG 212
Db 183 GAGVAPKAGVAGAFAGAPGVGPFPGGQPGVLPVAPKPLPAGYLPYKTKLPYFG 242
Qy 213 SGTSGRGLGCGAGAG-----AAAAAAGAGAGGCGYGLG-----GLGSGTSGR 257

```



```

Db 243 PGGAAGSAGKAGGPTGTGVPQAAAAAATAKAGAGAGVLPVGVGPGTIPGAPGAIIP 302
OY 258 --GGGCGGAGAAAAAAGAGAGGCGGCGTSGRG--GLGCGGAGAAAAAAGAA 313
Db 303 GGGGAGAGGAGDAAAAAATAKAFGAAGGIPGVGVPVGVGVPV----- 355
OY 314 AGGAGGCGT--GLGSGCTSGPGCGTSGCGTSGCGGAGAAAAAAGAA--GGAGC 369
Db 356 --GVGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 413
OY 370 GGGGGLSGGCGTSGRGGLGGG--AGAAAAAAGAGAGGCGGCGTSG--GGTGS 419
Db 414 VGGIGTPTFFGL-GPGGPGTIGDAAAAAATAKATGAGGAGLGGVPGVPGVPGVPGV 472
OY 420 GGGGCGGAGAAAAAAG--GAGCG-----GYGGLSGGCGTSG--GRG 460
Db 473 GVGCGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 532
OY 461 GLGCGGAGAAAAAAG--GGAG--GGGCGGCGTSGPG--GYGCGGCGTSGRG--G 511
Db 533 GYVAGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 591
OY 512 LGGGCG--GAAAAAAGAGAGGCGGCGTSGRGGLGGGAGAAAAAAGAAAGGA 569
Db 592 VPGGAVPGTTLAAAKAARFGPGGAGLGGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 650
OY 570 GGGGCGGCGTSGRGGLGGG--GAGAAAAAAGAGAGGCGGCGTSGRGGLGGG 625
Db 651 AOGGGLGGVGGCGLGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 709
OY 626 LGGGAGAAAAAAGAGAGGCGGCG--LGGGCGTSGPG 663
Db 710 AGGLG-----VGKPKPGGALGALG--PGG 735

RESULT 14
EBN1_EBV STANDARD: PRT: 641 AA.
AC P0311;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1998 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE EBNA-1 NUCLEAR PROTEIN.
GN BKRF1.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8427067; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Segun C.,
RA Tufnell P.S., Barrett B.G.;
RA "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RT Nature 310:207-211(1984).
RN [2]
RP SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE=86259739; PubMed=3460083;
RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RT "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RT proteins: a probable transcriptional initiation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=90266473; PubMed=2161150;
RA Petri L., Sample C., Kieff E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT latent infection nuclear proteins.";
RL Virology 176:563-574(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.

```

```

RX MEDLINE=96006523; PubMed=7553871;
RA Bokhariev A., Bartwell J.A., Pfuetzner R.A., Furey W.F. Jr.,
RA Edwards A.M., Traapier L.;
RT "Crystal structure of the DNA-binding domain of the Epstein-Barr
RT virus origin-binding protein EBNA 1.";
RL Cell 83:39-46(1995).
CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSCRIPTIONAL FACTOR
CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOLAR, SOMEWHAT
CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC WITH THE NUCLEAR MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: V01555; CAA24816.1; -.
DR EMBL: M13941; AAA45889.1; -.
DR PIR: A03773; Q0BE31.
DR PIR: S33021; S33021.
DR TRANSFAC: T00211; -.
DR PDB: 1VHT; 23-DEC-96.
KW Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW 3D-structure.
FT DOMAIN 87 352 GLY/ALA-RICH.
FT SEQUENCE 641 AA; 56427 MW; 4D161653E16FC341 CRC64;

Query Match 21.8%; Score 763; DB 1; Length 641;
Best Local Similarity 35.7%; Pred. No. 6.2e-28;
Matches 232; Conservative 22; Mismatches 209; Indels 186; Gaps 24;

OY 114 GAGGCGTGGGCG--TSGRGGLGGGAGAAAAAAGAGCGGCGTSGGPGGY 172
Db 5 GPGTGPNGGLGKBDTSPEGSGSGPQRRGDNHGRGRGRGRG--GRPGAPGSGS 62
OY 173 GPGGQTSGR-----GGLGGGAGAAAAAAGAGAGCGGCGTSGTSGRG 220
Db 63 GPRHMDVRRPKRPSCTGCKTHGTG--AGAGAGAGAGAGAGAGAGAG--GAGGAG 119
OY 221 GLGCGAGAAAAAAGAGAGCGGCGTSGRGGLGGGAGAAAAAAGAG 280
Db 120 GAGGAGAG-----GAGAGAG-----GAGAGAGAGAG--GAGAGGAGGAG 158
OY 281 GCGTGGGCGTSGRGGLGGGAGAAAAAAGAGAGCGGCGTSGGPGGY 340
Db 159 AG-----GAGGAGAGAG-----GAGAGAG-----GAGGAG-- 187
OY 341 QTSRGGLGGGAGAAAAAAGAGAGCGGCGTSGRGGLGGGAGAAAAAAG 400
Db 188 --GAGAGAGAGAG-----GAGAGAGAGAGAGAGAGAGAGAGAGAG 235
OY 401 AAGAGCGGCGTSGRGGLGGGAGAAAAAAGAGAGCGGCGTSGTSGRG 460
Db 236 GAGGAGAG-----AGAGAGAGAGAGAGAGAGAGAGAGAGAG--GAGGAGAG 282
OY 461 GLGCGGAGAAAAAAGAGAGCGGCGTSGGPGGYGCGGCGTSGRGGLGGGAGAA 520
Db 283 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
OY 521 AAAAAAGAGAGCGGCGTSGRGGLGGGAGAAAAAAGAGAGCGGCGTSG 579
Db 334 -----SGGRGCGSGGCGRGSGGRGRGRGRGRGRGRGRGRGRGRGRGRGR 386
OY 580 GGTG-----GR-----GGLGGGAGAAAAAAGAGAGAGAG 611
Db 387 GSSSSGPPRRPPRRPFFHVPGEADYFEYHOGGPDGPVPGVPGVPGVPGVPGVPGVPGV 446

```



```

OY 612 YGSLGSGTSGRGLG-----OGAGAAAAAAGG-----AG 646
DB 447 STGPRGCGDGRKKKGWFGKHKGGSNPKFENIAGLRALLARSVERTTEGTVAG 506
OY 647 QGGYGLGS-----OGTS-----GPGY--GPGQGTSGIR 674
DB 507 VFVYGGSKTSLYNLRCTALAIPOCRLTPLSLRPLFGMAPGPGPGLR 555

RESULT 15
ELS_CHICK STANDARD; PRT; 750 AA.
AC P07916;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
GN Gallus gallus (Chicken).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX Gallus.
NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87242320; PubMed=3593675;
RA Bressan G.M., Argos P., Stanley K.R.;
RT "Repeating structure of chick tropoelastin revealed by complementary
RT DNA cloning.";
RL Biochemistry 26:1497-1503(1987).
RN [2]
RP SEQUENCE OF 85-750 FROM N.A.
RX MEDLINE=88309083; PubMed=2841924;
RA Baule V.J., Foster J.A.;
RT "Multiple chick tropoelastin mRNAs.";
RL Biochem. Biophys. Res. Commun. 154:1054-1060(1988).
RN [3]
RP SEQUENCE OF 457-750 FROM N.A.
RX TISSUE=Aorta;
RL MEDLINE=87297534; PubMed=3502711;
RA Tokimitsu I., Tajima S., Nishikawa T., Fukasawa T.;
RT "Sequence analysis of elastin cDNA from chick aorta and
RT tissue-specific transcription of the elastin gene in developing chick
RT embryo.";
RL Arch. Biochem. Biophys. 256:455-461(1987).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NUCLEAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -1- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST TWO FORMS OF CHICKEN
CC ELASTIN THAT ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M18633; AAA48761.1; -
CC EMBL: M21880; AAA49082.1; -
CC EMBL: M15889; AAA49108.1; -
CC PIR: A27264; A27264.
CC PIR: A26601; A26601.
CC HSP: P04002; IWEA.
CC Structural protein; Connective tissue; Repeat; signal;
CC Alternative splicing.
FT NON_TER 1 24
FT SIGNAL <1 24
FT CHAIN 25 750 ELASTIN.

```

```

FT DOMAIN 83 686 8 X TANDEM REPEATS.
FT REPEAT 83 127 1.
FT REPEAT 219 262 2.
FT REPEAT 263 318 3.
FT REPEAT 319 393 4.
FT REPEAT 394 482 5.
FT REPEAT 483 554 6.
FT REPEAT 555 619 7.
FT REPEAT 620 686 8.
FT MOD_RES 63 66 8.
FT MOD_RES 66 66 8.
FT MOD_RES 111 111 1.
FT MOD_RES 111 111 1.
FT MOD_RES 115 115 1.
FT MOD_RES 156 156 1.
FT MOD_RES 159 159 1.
FT MOD_RES 198 198 1.
FT MOD_RES 200 200 1.
FT MOD_RES 235 235 1.
FT MOD_RES 252 252 1.
FT MOD_RES 256 256 1.
FT MOD_RES 297 297 1.
FT MOD_RES 301 301 1.
FT MOD_RES 354 354 1.
FT MOD_RES 357 357 1.
FT MOD_RES 427 427 1.
FT MOD_RES 431 431 1.
FT MOD_RES 513 513 1.
FT MOD_RES 517 517 1.
FT MOD_RES 520 520 1.
FT MOD_RES 586 586 1.
FT MOD_RES 590 590 1.
FT MOD_RES 593 593 1.
FT MOD_RES 655 655 1.
FT MOD_RES 658 658 1.
FT MOD_RES 719 719 1.
FT MOD_RES 721 721 1.
FT MOD_RES 743 743 1.
FT MOD_RES 748 748 1.
FT MOD_RES 750 750 1.
FT VARSPLIC 212 212 1.
FT VARSPLIC 501 501 1.
FT CONFLICT 536 536 1.
FT CONFLICT 571 571 1.
FT CONFLICT 610 610 1.
FT CONFLICT 654 654 1.
FT CONFLICT 667 667 1.
SQ SEQUENCE 750 AA; 63697 MW; E57ECD60C6E556F CRC64;

Query Match 20.2%; Score 704.5; DB 1; Length 750;
Best Local Similarity 36.7%; Pred. No. 2,4e-25;
Matches 284; Conservative 27; Mismatches 268; Indels 195; Gaps 43;

OY 15 SGRGGLG-GGAGAGAAAAAAGAGAGGCGTSGRGLG-GGAG-----GA 65
DB 45 AGVGGGLAGLGGAGAGGKPLKPGVSGIGPLGPGAGVGGAGLPGAPAPGA 104
OY 66 AAAAA--AAAAGAGGCGTSGRGL--GGGAGAAAAAAG-----113
DB 105 ASAAALAAAAAGAGLGGVGTG--GLGGVGGVPGGLGVPVGVGVGAAGAPKVP 162
OY 114 ---GAGGCGYGLGSGGT--SGRGLG--GGAGAAAAAAGAGGCGG-----GC 160
DB 163 AGIPGAPFGG-GVLPAGAGIPPGVGVLPVPTGTGIRAKGAGAGAFAGIPGGRLPVNG 221
OY 161 LGSQGTSGPGGPGGQGTSGRGLGCGAGAAAAAAG-GA-----GGGCGTGLG---212
DB 222 L-----GPGGIGAGVLAGKAGYPTGTGAGVGAQAAAAAAGAGAGVLPAGAGIPGVGV 275
OY 213 -SGTSGRGLGCGGAGAAAAAAGAGAGG--GGGGLG-----GGTSG 256
DB 276 PGVGVPGAGVGGPAAAAAAGAGAGATGAGVLPAGAGVPGVPGVPGVPLVPGVG 335

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 3, 2001, 15:05:16 ; Search time 95.41 Seconds
(without alignments)
944.343 Million cell updates/sec

Title: US-09-490-291-4

Sequence: 1 MGSNNHHHSGMSGRGL.....GGYGGQQTSGTRPAKLN 681

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5.

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_UNCLASSIFIED:*
13: SP_VERTEBRATE:*
14: SP_VIRUS:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1991.5	57.0	617	5	046172 nephila cla
2	1726.5	49.4	544	5	046171 nephila cla
3	1450	41.5	988	5	017434 nephila cla
4	1386.5	39.7	2639	5	076786 antherea p
5	1254.5	35.9	1489	2	053559 mycobacteri
6	1228.5	35.2	1079	2	053557 mycobacteri
7	1197	34.3	1306	2	053775 mycobacteri
8	1182.5	33.9	1381	2	053552 mycobacteri
9	1173	33.6	1660	5	053215 mycobacteri
10	1159.5	33.2	1884	5	09NHW2 nephila mad
11	1157	33.1	738	5	002402 plactada fu
12	1157	33.1	1538	5	053395 mycobacteri
13	1151	33.0	2249	5	09NHW4 nephila cla
14	1148	32.9	1468	5	09GUB5 galliera me
15	1145	32.8	1329	2	006810 mycobacteri
16	1139	32.6	853	2	053439 mycobacteri
17	1114.5	31.9	749	2	053844 mycobacteri
18	1099	31.5	907	5	044359 nephila cla
19	1083	31.0	871	5	044358 nephila cla

20	1075	30.8	882	2	053845 mycobacteri
21	1036	29.7	767	2	053435 mycobacteri
22	1024	29.3	837	2	053684 mycobacteri
23	1014	29.0	783	2	053809 mycobacteri
24	1011	28.9	636	5	016987 araneus dia
25	1004	28.7	731	2	050415 mycobacteri
26	984.5	28.2	714	2	053556 mycobacteri
27	984	28.2	741	2	006808 mycobacteri
28	975	27.9	810	11	09SES29 mus muscul
29	968.5	27.7	1011	2	P94586 mycobacteri
30	949	27.2	1729	5	09U617 dirosophila
31	923.5	26.4	923	2	053890 mycobacteri
32	922.5	26.4	667	2	053415 mycobacteri
33	914	26.2	626	5	09NHW1 nephila mad
34	903.5	25.9	694	2	053212 mycobacteri
35	896	25.7	904	5	076271 mytilus edu
36	893	25.6	484	2	053394 mycobacteri
37	856.5	24.5	610	5	09V5V8 dirosophila
38	846	24.2	922	5	044367 mytilus edu
39	839	24.0	902	5	016161 mytilus edu
40	833.5	23.9	576	2	P71664 mycobacteri
41	817.5	23.4	584	2	053575 mycobacteri
42	808.5	23.1	618	2	006801 mycobacteri
43	804	23.0	639	2	053952 mycobacteri
44	800	22.9	591	2	007224 mycobacteri
45	799	22.9	606	2	053884 mycobacteri

ALIGNMENTS

RESULT 1
ID 046172 PRELIMINARY; PRT; 617 AA.
AC 046172;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE DRAGLINE SILK PROTEIN SPIDROIN 1 (FRAGMENT).
OS Nephila clavipes (Ord spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID-6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94165058; PubMed-8120021.
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
RT proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus
RT bicentenarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Beckwith R., Arcidiacono S., Store R.;
RL Insect Biochem. Mol. Biol. 0:0-0(1998).
DR EMBL; U37520; AAC04504.1; -;
FT NON_TER 1 1
FT NON_TER 617 617
SQ SEQUENCE 617 AA; 49410 MW; 1EFOCE9269A832E2 CRC64;

Query Match 57.08; Score 1991.5; DB 5; Length 617;
Best Local Similarity 69.28; Pred. No. 2.1e-121;
Matches 457; Conservative 12; Mismatches 44; Indels 147; Gaps 29;

QY 11 GSMASRGGLGGGA---GAAAAAAGAGAGGCGYGLSGGSGRGGAGGAGAA 67
DB 7 GAGGCGYGLSGGCGRGGAGAGAGAGGCGYGLSGGCGAGGAGGAG--- 62
QY 68 AAAAAAGAGAGCGYGLSGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 127
DB 63 AAAAAAGAGAGCGYGLSGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 108

[illegible]

Query Match	49.4%	Score 1726.5	DB 5	Length 544
Best Local Similarity	63.3%	Pred. No. 2.1e-104	Mismatches 23	Indels 205
Matches 406	Conservative	7	Mismatches 23	Indels 205
Qy	16	GRGGLGCGAG-AAAAAAAAAAGAGAGGCGYGLSGSGTSGRGLGCGAGAAAAA	74	
Dy	1	GYGGLGCGAGGAGAGAAAAAAGAGGCGYGLSGSGTSGRGLGCGAGAAAAA	52	
Qy	75	GGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	134	
Dy	53	GGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	92	
Qy	135	GGGAGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	194	
Dy	93	GGGAGAGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	135	
Qy	195	AAAAAGAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	254	
Dy	136	AAAAAGAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	189	
Qy	255	SGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	314	
Dy	190	SGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	220	
Qy	315	GGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	374	
Dy	221	GGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	250	
Qy	375	LGSGGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	434	
Dy	251	LGSGGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	280	
Qy	435	AAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	494	
Dy	281	AAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	320	
Qy	495	SGPGYGPQGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	554	
Dy	321	SGPGYGPQGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	351	
Qy	555	AGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	614	
Dy	352	AGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	403	
Qy	615	LGSGGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	655	
Dy	404	LGSGGTSGRGLGCGAGAAAAAAGAGGCGYGLSGTSGRGLGCGAGAAAAA	437	
RESULT	3			
ID	017434	PRELIMINARY	PRT	988 AA.
AC	017434			
DC	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DE	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)			
DE	MINOR AMPULLATE SILK PROTEIN MISP1 (FRAGMENT).			
OS	Nephila clavipes (Orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OC	NCBI_TaxID=615.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98200471; PubMed=954138;			
RA	Colgin M.A., Lewis R.V.;			
RT	"Spider minor ampullate silk proteins contain new repetitive sequences			
RT	and highly conserved non-silk-like 'spacer regions'";			
RU	protein Sci. 7:667-672(1998).			
DR	EMBL: AF027735; AAC14589.1; .			
DR	InterPro: IPR000817; .			
DR	PRINTS; PR00341; PRION.			

	Query Match	49.4%	Score 1726.5;	DB 5;	Length 544;
	Best Local Similarity	63.3%	Pred. No. 2.1e-104;	Mismatches 23;	Indels 205; Gaps
	Matches 406;	Conservative			
Qy	16 GRGGLGCGAG--AAAAAAGAGAGGGYGLSGSGTSGRGLGCGAGAAAAAA	74			
Dd	1 GYGGLGCGAGGAGAAAAAAGAGGGYGLSGSQC---ACRGCGAG---AAAAA	52			
Qy	75 GGAGGGYGLSGTSGRGLGCGAGAAAAAAGAGGGCTYGLSGTSGRGL	134			
Dd	53 GGAGGGYGLSGSQC--ACRGGLGYGAG---AAAAGVGQ---GGL	92			
Qy	135 GGAGAGAAAAAAGAGGGYGLSGTSGPGYGPQQTSGRGLGCGAGAAAA	194			
Dd	93 GGAGAGAGAAAAAAGAGGGYGLSGSQC---AGRGSGGGAG---135				
Qy	195 AAAAAAGAGGGYGLSGTSGRGLGCGAGAAAAAAGAGGGCTYGLSGT	254			
Dd	136 AAAAAAGAGGGYGLSGSQC--ACRGGLGCGAG---AAAAAAGAGGGYGLGQ--189				
Qy	255 SGRLGCGAGAAAAAAGAGGGYGLSGTSGRGLGCGAGAAAAAA	314			
Dd	190 -----GAGGGYGLSGSQC--ACRGGLGCGAG---AAAAA	220			
Qy	315 GGAGGGYGLSGTSGPGYGPQQTSGRGLGCGAGAAAAAAGAGGGCTY	374			
Dd	221 GAGG-----GGLGCGAG---AAAAAAGAGGGYGLGQ--250				
Qy	375 LGSQGTSGRGLGCGAGAAAAAAGAGGGCTYGLSGTSGRGLGCGAGAAA	434			
Dd	251 LGSGC---AGRGCGAG---AAAAAAGAGGGYGLGQ-----280				
Qy	435 AAAAAAGAGGGYGLSGTSGRGLGCGAGAAAAAAGAGGGCTYGLSGT	494			
Dd	281 -----QGAGGGYGLSGSQC--ACRGGLGCGAG---AAAAAAGAGGQ-----320				
Qy	495 SGPYGPQQTSGRGLGCGAGAAAAAAGAGGGCTYGLSGTSGRGLGCG	554			
Dd	321 -----GGLGCGAG---AAAAAAGAGGGYGLGQGA-----GOG	351			
Qy	555 AGAAAAAAGAGGGYGLSGTSGRGLGCGAGAAAAAAGAGGGCTYGL	614			
Dd	352 AGAAAAAAGAGGGYGLSGSQC---ACRGCGAG---AAAAAAGAGGGYGL	403			
Qy	615 LGSQGTSGRGLGCGAGAAAAAAGAGGGYGLSGS	655			
Dd	404 LGGGY--GRGLGCGAGAAAV-----GAGGGYGLSGS	437			
RESULT	3				
ID	017434	PRELIMINARY; PRT; 988 AA.			
DC	017434				
RA	01-JAN-1998 (TrEMBLrel. 05, Created)				
RT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)				
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)				
DE	MINOR AMPULLATE SILK PROTEIN MISP1 (FRAGMENT).				
OS	Nephila clavipes (Orb spider).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;				
CC	Aranomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.				
OX	NCBI_TaxID=615;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98200471; PubMed=954138;				
RA	Coulgin M.A., Lewis R.V.;				
RT	"Spider minor ampullate silk proteins contain new repetitive sequences				
RU	and highly conserved non-silk-like 'spacer regions'";				
DR	EMBL: AF027735; AAC14589.1; -.				
DR	InterPro: IPR000817; -.				
DR	PRINTS; PR00341; PRION.				

RA Gordon S.V., Eiglmeter K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hohnsby T., Jagals K., Kiroh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL021246; CAA16067.1; -
 DR TubercuList: RV2490c; -
 DR InterPro: IPR000084; -
 DR InterPro: IPR000228; -
 DR InterPro: IPR002173; -
 DR Pfam: PF00934; PE; 1.
 DR ProDom: PD001223; -; 1.
 DR PROSITE: PS00583; PKB_KINASES.1; UNKNOWN_1.
 DR PROSITE: PS01287; RTC_UNKNOWN_1.
 SO SEQUENCE 1660 AA; 133124 MW; 3A889CE12C0FA945 CRC64;

Query Match 33.6%; Score 1173; DB 2; Length 1660;
 Best Local Similarity 35.7%; Pred. No. 2,2e-68;
 Matches 331; Conservative 36; Mismatches 281; Indels 280; Gaps 37;

QY 11 GSMASGRGLGGGAGAAAAAAGGA-----GCGGCG----- 46
 DB 451 GPVSGNGNGKGNCAHAPVACGCHGNGGAGNGGLVGDGAGHGCDAAAGVADMTAI 510
 QY 47 -LGSQGT-----SGRGGLGCG-GAGAAAAAAGAGAGCGGCGGCGG----- 88
 DB 511 FLGSSGTGTGDEGCGAGAGAGAGAGAHAGDAGAGAGGNGAGAGAGAHFNAVLVS 570
 QY 89 -----GTSRGGLGCGGAGAAAAAAGAGAGCGGCGG----- 123
 DB 571 GCGNGDGGAGRGCGDGGAG--GAGGDAFAGRAGSGCGVGDGAGAGAGAPNGSGCRDM 629
 QY 124 -----GSGGTSGRGGLGCGGA-----GAAAAAA 147
 DB 630 AFKDDGAGAGDGPAGCGAGKAGAGACATGATGATVHSGGNGKCGAGADATVAG 689
 QY 148 AAAGAGCGGTGGL-GSGGTSGPGGCG-----PG---QGTSGRG 182
 DB 690 ANGAGGAGGNGGLVGDGAGDGGGSAAGANGANGVGEDAGDTLSGPGEGSEANGGQG 749
 QY 183 GLGCGAGAAAAAAGAGAGAGCGGCGGGLGSGGTSGRGGLGCGGAGAAAAAAGAG 242
 DB 750 GVGCGGAGAGAGDGGAGSSALGSGGNGRGDAAG-----GAGGAGAGAGAGSVSGDGG 805
 QY 243 QCGYCGLGSGGTSGRGGLGCGGAGAAAAAAGAG-GAGGCGGLGSGGTSGRGGLGCG 301
 DB 806 PCGKCGAGAGAGAGASG-CGCGKAGSADSAEAVGAGAGCGKGGDGGV--GVCGDGGPGGD 862
 QY 302 G-AGAAAAAAGAGAGCGGCGGGL-GSGGTSGPGGCG-----PGQGTSGRG 347
 DB 863 GGAGGAGACAGCGSHGCGVCGDGLGAGGNGGDPGCHGSDGCGDGGDGG--AGGLG 920
 QY 348 LGG-OGAGAAAAAAGAG-----GAGAGCGGCGGLGSG-----GTSGR--GG 385
 DB 921 LGGSGNGCTRAASGVNADSDHPGSGGNGGNGAGNAGASVAGAGAGNGDGNAGVGGCG 980
 QY 386 LGGCGAGAAAAAAGAGAGCG-----GCGGGLGSG--TSGRGGLGCG-GAGAAAA 435
 DB 981 AGNGNGDCAAGANGANSAPGSDALALGCGNGGCGDAGAGAGAGAGAGAGAGAGAG 1040
 QY 436 AAAAAAGAGCGGCGGCGGCGG-----GGLGCGGAGAAAAAAGAGAGCGGCG 486
 DB 1041 DGGAGNGAGAGNGCGVAGAGAGANGIDSTIGTGGAGGCGGCGGAGVGGCHGDDGV 1100
 QY 487 GG-----LGSQGTSGPGGCG-----PGQGTSG 507
 DB 1101 GGAAAPSGTGVSHGTGCGVGDGGLGAGAGAGAGNGGIGITVCGAGAGAGCGGDPG--AG 1158

QY 508 GRGGLGCG-----GAGAAAAAAGAGAGCGGCGGCGG 543
 DB 1159 GGGGGLGCGSGNGCTSAANGVADSKHGRLTGGDGGVCGNGAKKAAAGDGGGDDG--GNAG 1216
 QY 544 TSGRGGLGCGGAGAAAAAAGAG-GGGGCG-----LGSQGTSGRGGLGCGG----- 593
 DB 1217 LFGDGGAGDGDGDTAAELAGDGGAGAGKGGAGDAGDIDGCGDGGKGGDGAHGALGLT 1276
 QY 594 -----GAAAAAAGAGAGAG-----GCGGCGGLGSGGTSGRGGLGCGGAGAAA-AAAAAAG 644
 DB 1277 VAGGNGAG 1336
 QY 645 AGGCGGGLGSGGTSGPGGCGGCGGCGG 672
 DB 1337 NGCGGTGCGGAG 1364

RESULT 10
 Q9NHW2 PRELIMINARY; PRT; 1884 AA.
 AC Q9NHW2;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
 GN FLAG.
 OS Nephila madagascariensis.
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 CC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
 OX NCBI_TaxID=115969;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20156766; PubMed=10688794;
 RA Hayashi C.Y., Lewis R.V.;
 RT "Molecular architecture and evolution of a modular spider silk protein
 RT gene."; 287:1477-1479(2000).
 RL Science; 287:1477-1479(2000).
 DR EMBL; AF218623; AAF36091.1; -
 DR InterPro: IPR000087; -
 FT NON-TER 1884
 SO SEQUENCE 1884 AA; 148738 MW; 27B6F45339FD20A5 CRC64;

Query Match 33.2%; Score 1159.5; DB 5; Length 1884;
 Best Local Similarity 39.7%; Pred. No. 1.8e-67;
 Matches 318; Conservative 34; Mismatches 296; Indels 153; Gaps 35;

QY 11 GSMASGRGLGGGAGAAAAAAGAGAGCGGCGGGLGSGGTSGRGGLGCGGAGAAAA 69
 DB 564 GGLGGRGPGGVPVPG-GSGPGSIGPGSGGCGGLGPGSGGCGGLGPGSGGPGVSGG 622
 QY 70 AAAAAAGAGGCGGGLGSG-OGTSGRGGLGCGGA-----GAAAAAAGAGAGAGCG 120
 DB 623 GPGGGSAGPGVGAAGGPGGSGGPGAGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 681
 QY 121 GGLGSGGTSGRGGLGCGGAGAAAAAAGAGAGAGCGGCGG-----GGLGSGGTSGRG 173
 DB 682 GPGGPGGAGAGGPGGPGGPGGAGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 741
 QY 174 PGQGTSGRGGLGCGGAG-AAAAAAGAGAGAGCGG--GGLGSGGTSGRGGLGCGGAG-A 229
 DB 742 PG-GAGPGGCGPGGAGPGGSGPGGIGPGGSGPGGIGPGGT-GPGGAGPGGAGG 798
 QY 220 AAAAAAAGAGAGGCGGGLGSG-OGTSGRGGLGCGGAGAAAAAAGAGAGAGAGAGAG 287
 DB 799 GAGPGGAGGAGGPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 858
 QY 288 GSG-----GTSRGGLGCGGAG-AAAAAAGAGAGAGCGG 321
 DB 859 GGTITIEDLDTVDANGPRTISEELTIGACAGAGVPGGSGPGGSGPGGSGPGGSGPGG 918
 QY 322 Y--GGLGSGGTSGPGGCGGPGGTSGRGGLGCGGAG-AAAAAAGAGAGAGCGGCGGGLGSG 378

```

Db 919 SGPBGVGS -GSGPGGVGSGGFGPGGCTGPGGSGPBGVGPBGVGP 975
QY 379 GTSRGGLGCGCQ-----AGAAAAAAGAGAGCGGCGGCGTSGRGL 424
Db 976 GSGSGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1034
QY 425 GGCGAGAAAAA-AAAAAGAGAGCGGTGGLGSGCTSGRGLGCGAG-----AAAAA 477
Db 1035 GCEGPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1090
QY 478 AGGAGCGGTGGLGSGCTSGPBGVGPBGVGPBGVGPBGVGPBGVGPBGV 528
Db 1091 PGAGAGPGGAG-----PGGPGGAGAGAGAGAGAGAGAGAGAGAGAG 1146
QY 529 GAGAGCGGTGGLGSGCTSGAGGCGGAG-----AAAAAAGAGAGAGAG 575
Db 1147 GSGSGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
QY 576 --GLGSGGTS--GRGGLGCGAG-AAAAAAGAGAGAGAGAGAGAGAGAG 629
Db 1207 GAGPGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
QY 630 GAG-----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 655
Db 1267 GSGGTTTIEDITDADPTTISELTISGAGAGSGGPGAGPGVGPBGV 1326
QY 656 GGTSGPGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 676
Db 1327 PGSGPGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1346

```

RESULT 11

```

002402 PRELIMINARY: PRT; 738 AA.
ID 002402:
AC 002402:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE INSOLUBLE PROTEIN.
OS Pinctada fucata.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidea;
OC Pterioidea; Pteritidae; Pinctada.
OX NCBI_TaxID=50426;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97320490; PubMed-9177341;
RX Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,
RA Nakashima K., Takahashi T.;
RT Structures of mollusc shell framework proteins."
RL Nature 387:563-564(1997).
DR EMBL; D86074; BAA20466.1;
SQ SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;

```

Query Match 33.1%; Score 1157; DB 5; Length 738;
 Best Local Similarity 43.8%; Pred. No. 1.2e-67;
 Matches 330; Conservative 30; Mismatches 191; Indels 202; Gaps 27;

```

QY 14 ASGRGGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 52
Db 61 SAGLYGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 53 SGRGGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 92
Db 121 GAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 93 -----RGLGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145
Db 181 DLGAALALGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 146 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 203

```

```

Db 241 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 294
QY 204 GCGGTGGLGCGTSGRGLGCGAG-----AAAAAAGAGAGAGAGAGAGAG 259
Db 295 GGGGGLGCGGTG--GRGGRGGRGRRRAAAAAAAGAGAGAGAGAGAGAG 344
QY 260 LGGGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295
Db 345 GGGGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
QY 296 -----GGLGCGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAG 446
Db 405 QIDDLKDVLTDLAAGLLKSSASASASASASASAGCGGCGGNGC----- 446
QY 343 SGRGGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402
Db 447 -GNGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
QY 403 GAGAGCGGTGGLGSGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAG 462
Db 489 -GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
QY 463 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
Db 536 RGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585
QY 523 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 581
Db 586 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 628
QY 582 TSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634
Db 629 -----GGSADVAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 678
QY 635 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
Db 679 -GSGGSGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 710

```

RESULT 12

```

005395 PRELIMINARY: PRT; 1538 AA.
ID 005395:
AC 005395:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PGRS-FAMILY PROTEIN.
GN PE_PGRS OR RV3345C OR MYV004.01C-MTV016.45C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE-98295987; PubMed-9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulton J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Parkhill J.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021841; CAAl7117.1;

```

DR HSSP; P00441; 1SOS.
 DR Tuberculist; RV3345C; -
 DR InterPro; IPR000084; -
 DR InterPro; IPR002173; -
 DR Pfam; PF00934; PE; 1.
 DR ProDom; PD001223; -; 1.
 DR PROSITE; PS00583; PFKA_KINASES_1; UNKNOWN_3.
 SQ SEQUENCE 1538 AA; 129386 MW; 788F0B205587592 CRC64;

Query Match 33.1%; Score 1157; DB 2; Length 1538;
 Best Local Similarity 34.4%; Pred. No. 2, 2e-67;
 Matches 317; Conservative 36; Mismatches 308; Indels 260; Gaps 31;

QY 1 MRGSHHHHHHSMASGRCGLGCGAGAAAAAAGG-----ACGGYGLGSGGTS 53
 DB 549 LMSAGANGTAPTSGNGGNGAGATPTVAGENGAGNGHGSGVNGAGAGANGVA 608
 QY 54 G-----RGGLGGGA-GAAAAAAGAGAGGGGGLGSG-----GTS 91
 DB 609 GTGLALNGNGNGGICGNGGSAAGTGDDGKGNGAGANGQDFSAANGANGGCGNG 668
 QY 92 GRGGLGCGAGAAAAAAGAGAGCGCGYGLGSGTSGRGLGCGA-----139
 DB 669 GNGIGCGGDAFATFAKAGNGAGNG-----GNVVAAGCGAGAGKALPAMKATGADG 724
 QY 140 -----GAAAAAAGAGAGCGGCG-LSGSGTSGPGYGPQG-----T 178
 DB 725 TAPTSGDNGNGNGASPTVAGNGGCGKSGGNGNGGAGNGAGNGAQAQAGTGPPT 784
 QY 179 SGRGGLGCGAGAAAAAAGG--AGCGYGLGSGTSGRGLGCGAGAAAAA 236
 DB 785 SGDSGISTDGGAGNGAGAGGTLAGHGNGSGK--GNGCGTIGAGRGADGAPN 842
 QY 237 AAGGAGCGYGLGSGTSGRGLGCG--GAGAAAAAAGAGAGCGGCG-----286
 DB 843 ANGANGENG-----GSGGNGDDGAGNGAGAGKQAAGYTDGATGTGGDNGDGGKAG 898
 QY 287 -----LSGSGTSGRGLGCG-----301
 DB 899 DGGAGENGLSGAMLPGCGTYVGNPTGCGNGGNGAGVGTGCKAGTGLTGLDGTGTT 958
 QY 302 -----GAGAAAAAAGG-----AGCGYGLG-----LSGSGTS 331
 DB 959 PNGNGNGNGNGAGTAGNGSGAGNGNGSGGLNGDNGNGGAGALNQAFFGTG 1018
 QY 332 GPGGY-----GPGQOTSGRGLGCGAGAAAAAAGAGAGCGYGLGSGTSGRGLG 387
 DB 1019 GKGGNGNGAGMINGLGGCGGAGAVDYATTGGAGNGAGCGFASRGLGPGAG 1078
 QY 388 GCGAGAAAAAAGAG-----GCGYGLGSG-----TSRGLGCGAGAAAAA 436
 DB 1079 GPGAGCGFASGVGAGAGDGGAGGCGGCGGCGGEGRTGNGSGCGCGGCGISLGG 1138
 QY 437 AAAAG-----AGCGYGLGSGTSGRGLGCGAGAAAAAAGAGAGCG 484
 DB 1139 NGGLAGNGGVSETFEGAGAGNGTGGPG--GPEGNGGLGAGGAGNGGVSTTGGDAG 1196
 QY 485 GYGGLGSGTSGRGLGCGAGAAAAAAGG-----AG 532
 DB 1197 GKGGNGDG--GNVGLGDMAGSGAGNGGIGTIDAGAGAGAGAGNGSGSKTTTGNAG 1254
 QY 533 GCGYGLGSGTSGRGLGCGAGAAAAAAGG-----AAAGAGCGYGLGSGGTS-GRGL 588
 DB 1255 SCGGGNGCGTGLNAGAGAGAGNAAGVAFSGNAVGGDNGNGGNGGCGTGGAGGK 1314
 QY 589 GCGAGAAAAAAG-----AAAGAGAGCGYGLGSGTSGRGL-----GCGAGAAA- 635
 DB 1315 GGNSSGAAGSGVYNTTACHGNGGNGNGAGAGAGCGGAGSAGNGHGGATGG 1374
 QY 636 -----AAAAAAG-----AGCGYGLGSGGTS-----G 660
 DB 1375 DCGNGGNGSGNSTYVAGLAGAAGAGNGCGTSSAAAGHGCGSGSGSTTTGAGAAAG 1434

QY 661 PGYGP-----QOTSGIRR 675
 DB 1435 NGAGAGGGLSTGSGGPRR 1455

RESULT 13
 Q9NHM4 PRELIMINARY; PRT: 2249 AA.
 AC Q9NHM4;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-Mar-2001 (Tremblrel. 16, Last annotation update)
 DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
 GN FLAG.
 OS Nephila clavipes (orb spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
 OX NCBI_Taxid:6915;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-20156766; Pubmed-10688794;
 RX Hayashi C.Y., Lewis R.V.;
 RT "Molecular architecture and evolution of a modular spider silk protein gene."
 RL Science 287:1477-1479(2000).
 DR EMBL; AF218621; AAF36090.1; -.
 DR InterPro; IPR000087; -.
 DR InterPro; IPR00209; -.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
 FT NON_TER 1 1
 FT 2249 2249
 SQ SEQUENCE 2249 AA; 174871 MW; 8BC8B168A147CDA0 CRC64;

Query Match 33.0%; Score 1151; DB 5; Length 2249;
 Best Local Similarity 36.7%; Pred. No. 7, 5e-67;
 Matches 330; Conservative 30; Mismatches 288; Indels 250; Gaps 36;

QY 10 HGSNAGRGRLGCGAG-----AAAAAAGAGAGCGG-----GLGSGGT 52
 DB 166 YPGGAGPGYGPAGAGPGYGPAGAGPGYGPAGAGPGYGPAGAGPGYGPAG 225
 QY 53 -----SGRGLGCGAG-----AAAAAAGAGAGCGG-----GLSGGTSGR 94
 DB 226 VPGGAGPGYGPAGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAG 285
 QY 95 -----GLGCGAGAAAAAAG-----AAAGAGCGYGLGSGTSGRGLGCGAGAAA 143
 DB 286 AGPGAGLGGAGPGAGTSGAGPGAGPGAGAGGAGGAGRGAGRGAGRGAGRG 345
 QY 144 AAAAAAGAGAGCGYGLGSG-----GTS 170
 DB 346 AGRGAGAGAGAGAGAGAGGSGGTTTIEDLITIDGADPITISBELTIGAGAGSGPG 405
 QY 171 GYGPQOTSGRGLGCGAGAAAAAAGAGAGCG--GYGLGSGT-----SGR 219
 DB 406 GAGPGNVGPRSGPGYGP-----GSGPGYGPAGPGYGPAGPGYGPAGPGY 459
 QY 220 GGLGCGAGAAAAAAGAGAGCGYGLGSGTSGRGLGCGA-----265
 DB 460 GGVGTGVPAGAGPGYGPAGPGAGAGSAG--GTYGAGGPGAGPGAGPGAGPG 517
 QY 266 -----GAAAAAAGAGAGAGCGYGLGSGTSGRGLGCG--GAGAAAAAAGAGAG 318
 DB 518 GAGPGYGPAGPGYGPAGAG--GPGYGPAGAGPGYGPAGAGSYLGAGAGSGGPGSG 576
 QY 319 GGY--GGLG-----SGTSGPGYGPAGQOTSGRGLGCGAGAA-----356
 DB 577 PGYGPAGAGPGYGPAGPGSGPGYGPAGPGSGGCGYGPAGPGSGPGYGPAGTGP 636
 QY 357 -----AAAAAAGAGAGCGYGLG-----GTSRGLGCGAG--AAA 396

```
Db 637 SGGYPCGSGPCGSGPCGSGPCGSGPCGSGPSTVPGCGSPGCGSPGCGAG 696
Qy 397 AAAAAAGAGGCGY-GLSGSGTSGRGLGCGAG-AAAAAAAAAGAGGCGYGLGSG 454
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 697 PGAGPGGAGPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 756
Qy 455 GTSRGLGCGG-GAGAA-----AAAAAAAAAGAGGCG 485
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 757 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 816
Qy 486 YGGLAS-OGTSGPGYGPQOQTSRGLGCGAGAA-----AAAAAAAAAGAG- 531
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 817 VCPGSGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG 876
Qy 532 GCGG-YGGLGSGGTSGRGLGCGAGAA-----AAAAAAAAAGAG- 569
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 877 GPGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG 936
Qy 570 -----GCGGTCG-LGSGGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAG 619
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 937 PGVYGPAGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG 996
Qy 620 TSGRG-----GLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 997 PGGSGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG 1048
```

```
RESULT 14
OGGUB5 PRELIMINARY: PRT: 1468 AA.
AC OGGUB5.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HEAVY-CHAIN FIBROIN (FRAGMENT).
GN FIB-H.
OS Galleria mellonella (Max moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditytisia;
OC Pyraloidea; Pyralidae; Galleriinae; Galleria.
OC NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POSTERIOR SILK GLAND;
RA Zurevec M.; Kodrik D.; Yang C.; Sehnael F.;
RT "Heavy-chain fibroin of Galleria mellonella L.";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF095239; MAG10393.1; -.
FT NON_TER 1468 1468
FT SEQUENCE 1468 AA; 122705 MW; 2DA59E1181BB3DDE CRC64;
```

Query Match 32.9%: Score 1148; DB 5; Length 1468;
Best Local Similarity 30.1%: Pred. No. 8, 2e-67;
Matches 354; Conservative 78; Mismatches 215; Indels 528; Gaps 36;

```
Qy 14 ASGRGGLGCGA-----GAAAAAAAAAGAGAG----- 41
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 267 ASGLGLGAGAGSAGSAGAGLGVGVGSSSSAASAGAGEVLLIDRSSAA 326
Qy 42 -----GGYGLGSGGTSGRGLGCGAG-- 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 327 AAGSAGSAGVGLGLGGLPGLGIGIPATSTSGAGLGLGVGAGAGAGAG 386
Qy 64 -----GAAAAAAAAAGAGAG-----GGYGLG 86
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 387 ASAGAGLGLGAGGSSGSSAASAGAGAGEVYIDRSSAAASAGSAGLGLG 446
Qy 87 SGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 140
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 447 LGLGLGPGGIGLNGVSASALGAGLGVGTAGASGLGAGAGVAVPAGAGLGVGAG 506
Qy 141 AA-----AAAAAAAAAGAGAGGCGYGLGSGGTSGP-GGYGPG 175
```

```
Db 507 GSSGSSAASASARGPVYIEDGSSAASAAAAGSAGLGLG-LGAMPPLGIGRN 565
Qy 176 QOTS-----GRGGLGCGAGAGAA-----AAAAAGAG 204
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 566 EVSASATGSAAGTGAAGLGGAGGAGSSAASAGAPVYIEDGSSAASAGAG 625
Qy 205 QCGYGLGSGGTSGRGLGCGG-----AG 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 626 ASGLGLGIGAMGTGLGIGTPNGVSSASATGSAAGTSGGLCGSAGAGSSLASAGAGAA 685
Qy 230 A-----AAAAAG-----GAGGCGYGLGSG 252
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 686 APVYIYIEDGSSAASAAAGSAGSVGLGLGLALPLGLGIPGASSAGSAGLGVGAA 745
Qy 253 GTSRGLGCGAG-----GAAAAAAAAAGAGAG----- 281
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 746 GTSGLGIGTGVASTAGSAGAGLGLGAGGSSGSSAASAGSAGAGEVYIDRSSAAS 805
Qy 282 -----GGYGLGSGGTSGRGLGCGAG 304
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 806 AAGSAGSAGPGLGLGVWPLGLGIPGASSASAGAGLGVGAAGTSGLGLGAGAG 865
Qy 305 A-----AAAAAGAGAGAG----- 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 866 AAGSAGAGLGLGAGSSGSSAASAGTSGAGEVYIDRSSAASAAAGSAGLGL 925
Qy 322 -----YGLGSGGTS-----GPGYGPQOQTSRGLGCGAGAGAGAGAGAG 367
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 926 GLGLGLPYGIGLGNVSSASALGAGLGVG-----TAGAGLGLGLGTTAGSAGAGAG 981
Qy 368 GCGGTCGLGSGGTSGRGLGCGG-----AGAGAGAGAGAGAGAGAGAGAGAG 412
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 982 GYAGAGSFGSSAASAGSAGAGEVYIDRSSAASAAAGSAGSGPLGLGLGVWPLG 1041
Qy 413 LGSQTS-----GRGGLGCGAGAGAA-----AAAA 440
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1042 IGNGVSSASATGSAAGTGAAGLGGAGSAGSAGSAGAPVYIEDGSSAASAA 1101
Qy 441 GAGAGGCGYGLG-----SGTSGRGLGCGAG----- 467
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1102 AGSAGLGLGLGAMGPLGIGTPNGVSSASATGSAAGTGAAGLGGAGSAGSAGS 1160
Qy 468 GAA-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1161 GAAGPAPVYIEDGSSAASAGSAGLGLG-LGAMGPLGIGTPNGVSSASATGSA 1219
Qy 503 GQOTSGRGLGCGAGAGAG-----AAAAAGAGAGAGAGAGAGAGAGAGAG 544
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1220 GSTGAGLGGAGSSAASAGAGAPVYIYIEDGSSAASAGSAGVGLGLSL 1279
Qy 545 SGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 596
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1280 GPLGIGTPHGVSSASALGAGLGVGAPGASGLGLGVNAGSAGAGLGLGAGGSSG 1339
Qy 597 -----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1340 LSTASAGSAGAGEVYIYIDRSSAASAGSAGLGLGLGIPGIGLGNVSSA 1399
Qy 635 AA-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1400 SALGAGLGVGTAGSAGLGLGTGASAGAGAG 1434
```

RESULT 15
006810 PRELIMINARY: PRT: 1329 AA.
AC 006810;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHEICAL 107.4 KDA PROTEIN.
GN PRGS-FAMILY OR RV1450C OR MTCY493.04.

OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP
RP SEQUENCE FROM N.A.
RC
RC SRR141-437RV;
RX MEDLINE=9825987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RX Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RX Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RX Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RX Hornsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RX Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RX Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RL EMBL; Z95844; CAB09271.1; -.
DR TuberculList; RV1450C; -.
DR InterPro: IPR000084; -.
DR InterPro: IPR002173; -.
DR InterPro: IPR003536; -.
DR Pfam: PF00934; PE: 1.
DR PRINTS; PRO1370; TRNSINTIMINR.
DR PRODOM; PD001223; ?; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
KW Hypothetical protein.
SQ SEQUENCE 1329 AA; 107410 MW; 777A125F6DBAB334 CRC64;

Query Match	32.8%;	Score 1145;	DB 2;	Length 1329;
Best Local Similarity	35.9%;	Pred. No. 1.2e-66;		
Matches 323;	Conservative 35;	Mismatches 300;	Indels 242;	Gaps 35

OY	11	GSMSAGRGGLGCGQA-----GAIAAAAAAAAAAGGA-----GCGGYG	45
Db	139	GGTIMGKMGAGGSGAPQGVGAGGAGLFGTGGGAGGAGAGAGGAGCGGMLTNGVg	1989
OY	46	GLGSGCTSGRGGLGCGGAGAAAAAAGAGAGCGGCGTGLGSGCTSGRGGLGCGGAGAAA	1059
Db	199	GAGGQSLLG-CATGGACGMAGLFGVGGTGGRCRCRCGCGYg--GTGAGGLGGTLYAGG	2555
OY	106	AAAAAAGG-AGCGGYGLGSGQ---GTSGRGLGEGGAGAAAAA-----AAAA	1500
Db	256	HGCGAGCGPCPTGCGVGHGCGYGAAGLLTGVGGHCGAGGAGGAVAGAAEDLPHOTSGYg	3151
OY	151	GGAGGCGYGLGSL--GCTSGRCGCTGPPQCTSGRGGLG-----GGAGAAAAA	2000
Db	316	GDAADGCTGGCGGMLAAGGAGGAGGCGTGGAGAGAFSRALTVAGDNGDPCGAGAGGT	3759
OY	201	GGAGG--GGYRGGLSGGCTG-CRGGLGGGAGAAAAAAGGA-----GGGCGYGLG	2500
Db	376	GGAGSTTGAHGAAGASPTSGGNGGAGGNGAHFSSGGAGAGNGGAGGAGGLVGNAGGAGAG	4355
OY	251	SGCT-----SGRGGLGCGGAGAAAAAAGGA-GCGGYGLGSGCTSGR	2959
Db	436	GNGAPGAPPSGCGDPPNGGGGAGGAGCGGCGGAGQADGAGAGAGGCGGNGATGATGL	4959
OY	296	GGL-----GGGAGAAAAAAGGA-----AGGA-----GCGGT	3422
Db	496	NGTLAGADGTDGKGKNGAGAGGGGAGGAGGCGKLAATTHODGSMGACGAGGNGGAGGKGID	5555
OY	323	GGGLSGGCT-----SGPGGCGGCGCTSGRGGLGCGGA-----	3533
Db	556	GGNGAKGCTPDMGCGDYGNGGNGGSGKRTGGAGGCTGGAGSTRAGADGANGATPTSGGNGCTG	6151
OY	354	--GAAAAAAGAGAGAGCGGCGT-GSGCTSGRGGLGCGGAGAAAAAAGAGAGC--	4070
Db	616	GNGANATVYAGGAGAGGRCGCGNGLVNGAGAGCGKGDMDAGVAGSPTTADESCTSGGNGC	6755
OY	408	-GGYGLGSGQ-----GTSGRGLGCGGAGAAAAAAGG-----AGCGGY	4489

```

Db      676 AGGAGGAGCGRGDPTGGGGGTGAGGAGNANANTTTTCAGCGDGGCHGCPAAGGNGGGGP 735
QY      449 GGL-----GSGGTSGRGGLGGGAGAAAAAAAAGAGGCGGTG---GLGS 491
Db      736 GGLGAGLFLGQNGIQGVGSGGKAGAGGLADGGNGANGNPFACDGGCHGNGGNPAGG 795
QY      492 QGTSGPGGYGPGQQ-----TSGRGGLGGGAGAAAAAAAAGAGGCGGTGGL-SSQ 542
Db      796 QGGSGGAGSTTGAAGAGHFTPTSGDGGDGGNGCN-----SDVGGNGGDDGGNGNGGSA 850
QY      543 GTSRGGGLGGCGAGAAAAAAA-----AAAGAGAGCGGCGGLGSGGTSGRG 587
Db      851 GTGGNGGCGGDDGACGMSANNTNPENGEMPCNGNCGGAGGAGAGAGLNG-GAGGAGGNG 909
QY      588 LGGGCGAAAAAAAAGAGGQ---GGYGGGLG-----SGGTSGRGGLGG 629
Db      910 LGGCGGCGAGGANGVAAGCAPGCGPGGAGHGAGGAGGAGGAGGCGGCVSPDGAGGAGAGGD 969
QY      630 G-----AGAAAAAAAAGAGGCGGCGYGC-LGSGGTSGPGGYGPGGQGTSTIRPAKL 680
Db      970 GGAPEDGANGNGCGACAFAGGGGGRGGDGGAGNAGAGPGGTG---STAKACPASTI 1028

```

Search completed: July 3, 2001, 15:05:26
Job time: 591 sec

xx Claim 1; Column 29-34; 68pp; English.
ps
xx The present sequence represents a spider silk protein from the present
cc invention. Spider silk proteins, and peptide fragments within the
cc proteins, can be produced and purified independently and can then be
cc mixed and made into fibres that have higher tensile strengths and
cc elasticity than naturally occurring fibres. The fibres can be used in
cc mixed composites. The invention allows the two naturally occurring
cc Nephila clavipes silk proteins to be produced independently so that
cc they can later be combined to form silk fibres of high tensile strength
cc and elasticity.
xx
xx
SQ Sequence 718 AA:

Query Match 60.3%; Score 2138; DB 19; Length 718;
Best Local Similarity 66.3%; Pred. No. 1.4e-145;
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;

QY 25 OGAGAAAAAAGAGAGCGGCGTSGRGLGCGAG-AAAAAAAAAGAGC 81
DB 1 qgaq-----aaaaaagagagagagagagagagagagagagagagag 56
QY 82 GGYGGLGSGGTSGRGLGCGAGAAAAAAGAGCGGCGTSGRGLGCGAG 141
DB 57 ggygglgagag-----agrggagag-----aaaaaagagagagagag 107
QY 142 GAAAAAAGAGAGAGCGGCGTSGTSGPGCGTSGRGLGCGAGAAAAA 201
DB 108 g-----aaaaaagagagagagagagagagagagagagagagag 142
QY 202 AGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAG 252
DB 143 ag 201
QY 253 SGTGSRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
DB 202 agg-agrgglgagagagagagagagagagagagagagagagagag 260
QY 292 GGTSGRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350
DB 261 gq-----agrggagag-----aaaaaagagagagagagagag 311
QY 351 GGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 392
DB 312 gggag 370
QY 393 AGAAAAAAG 452
DB 371 ag-----avaaaaagagagagagagagagagagagagagagag 419
QY 453 LGSOGTSGRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 499
DB 420 lgnqg-agrgglgagagag-----aaaaaagagagagagagagag 475
QY 500 GCGYCP-GOOTSGRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 552
DB 476 ggygglgagagag-----gggagagagagagagagagagagagag 531
QY 553 GGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592
DB 532 gggag 587
QY 593 OGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
DB 588 ggaq-----aaaaaagagagagagagagagagagagagagagag 636
QY 653 GGLGSGTSGPGYGPQOTSGRIRAPST 683
DB 637 ggygagagagag-----saaarllspqas 659

RESULT 2

AAV59070 ID AAV59070 standard; Protein; 718 AA.
xx
xx AAV59070;
AC
xx 08-MAR-2000 (first entry)
DT
xx
xx N. clavipes spider silk protein 1.
DE
xx
xx Spider silk protein; dragline silk protein; major ampullate gland; fiber.
KW
xx
xx Nephila clavipes.
OS
xx
xx US5989894-A.
PN
xx
xx 23-NOV-1999.
PD
xx
xx 04-OCT-1994; 94US-0317844.
PF
xx
xx 15-APR-1991; 91US-0684819.
PR
xx 20-APR-1990; 90US-0511792.
XX
xx
xx (UYWY-) UNIV WYOMING.
PA

PI Hlman MB, Xu M, Lewis RV;
XX
XX WPI; 2000-061225/05.
XX
XX N-PSDB; AA238195.
DR
XX

PT Isolated DNA, vector and transformed cell encoding for and useful in
the production of spider silk protein -

Claim 1; Fig 6A-D; 65pp; English.

CC The invention provides isolated cDNA molecules coding for spider silk
CC proteins. The spider silk proteins are characterized by repeating alpha
CC and beta regions and optional variable regions. The DNA sequences are
CC useful in the production of spider silk protein by recombinant DNA
CC techniques. The recombinant spider silk proteins may be used for the
CC production of fibers. The present sequence represents the spider silk
CC protein 1, derived from the major ampullate gland of Nephila clavipes.
XX
XX

SQ Sequence 718 AA:

Query Match 60.3%; Score 2138; DB 21; Length 718;
Best Local Similarity 66.3%; Pred. No. 1.4e-145;
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;

QY 25 OGAGAAAAAAGAGAGCGGCGTSGRGLGCGAG-AAAAAAAAAGAGC 81
DB 1 qgaq-----aaaaaagagagagagagagagagagagagagagagag 56
QY 82 GGYGGLGSGGTSGRGLGCGAGAAAAAAGAGCGGCGTSGRGLGCGAG 141
DB 57 ggygglgagag-----agrggagag-----aaaaaagagagagagag 107
QY 142 GAAAAAAGAGAGAGCGGCGTSGTSGPGCGTSGRGLGCGAGAAAAA 201
DB 108 g-----aaaaaagagagagagagagagagagagagagagagag 142
QY 202 AGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAG 252
DB 143 aggag 201
QY 253 SGTGSRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
DB 202 sgg-agrgglgagagagagagagagagagagagagagagagagag 260
QY 292 GGTSGRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 350
DB 261 gq-----agrggagag-----aaaaaagagagagagagagag 311

protein superfibre; major ampullate silk; orb web spider.

Nephila clavipes.

EP452925-A.

23-OCT-1991.

18-APR-1991; 91EP-0106217.

20-APR-1990; 90US-0511792.

(UTWY-) UNIV OF WYOMING.

Lewis RV, Xu M, Hinman M;

WPI: 1991-312199/43.

N-PSDB; AA014183.

DNA encoding spider silk protein-1 and 2 and variants - Isolated from Nephila clavipes, for prodn. of spider silk protein and fibres having desired characteristics

Claim 15: Page 23; 48pp; English.

The spider silk protein contains a basic 34 amino acid repeat. The repeat itself contains 3 regions. The first comprises 0-9 amino acids with a sequence AGR(GGX)2. This region is not highly conserved. The second region has a sequence GAG(A)X which is highly conserved and is 8-10 amino acids long. The third segment is (GGX)5 and is 15 amino acids long and is very highly conserved. In most cases X is A,Q, Y or L. Removal of the poly-(Ala) segments results in a silk having lower elasticity.

Sequence 718 AA:

Query Match 60.1%; Score 2129; DB 12; Length 718;
Best Local Similarity 66.4%; Pred. No. 6.2e-145;
Matches 495; Conservative 21; Mismatches 57; Indels 172; Gaps 35;

25 QGAGAAAAAGAGAGGCGGCGGCGT--SGRGLGCGAG-AAAAAGAGAGAG 81
||||| ||||||| ||||||| ||| | ||||||| ||||||| ||||||| |||||||
1 1ggag----aaaaaaggagggggggg1gggaggggggg1gggaggaagaaagagag 56
82 GGYGCGGCGTSGRGLGCGAGAAAAAGAGAGGCGGCGTSGRGLGCGAG 141
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
57 gggggggag----agrgggag----aaaaaaggaggggggg1gggag 107
142 GAAAAAGAGAGGCGGCGGCGT-----SGPGGYGP-GQQTSGRGL 186
| ||||||| ||||||| ||| | : ||||| | : |||||
108 g---aaaaaaggaggggggg1gggagggagggagggaggg1gggag 164
187 GGGGAGAAAAAGAGAGGCGGCGTSGRGLGCGAGAAAAAGAGAG 237
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
165 ggggagaaaaaaggaggggggg1gggaggggggg1gggag--aaaa 220
238 AAGAGAGGCGGCGGCGTSGRGLGCGAGAAAAAGAGAGGCGGCGTSGR 297
||||| ||||||| ||| | : ||||||| ||||||| ||||||| |||||||
221 aaagagag-----gg1gggagggaggaagaaagaggggg1gggag---- 262
298 GGGGAGAGAGAGAGAGGCGGCGTSGPGGYGP-GQQTSGRGLGCGAG 356
| ||||| ||||||| ||||||| ||| | : ||||| | : ||||||| |||||||
263 agggagag----aaaaaaggaggggggg1gggag--aggggggg1gggag 317
357 AAAA-----AAAAAGAGAGGCGGCGGCGTSGRGLGCGAGAAAA 398
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
318 aaagaggggg1gggagggaggaagaaaggggg1gggag--agrggg1gggag--a 373
399 AAAAAAGAGAGGCGGCGTSGRGLGCGAGAAAAAGAGAGGCGGCGTSGR 458
||||| ||||||| ||||||| ||| | : ||||||| ||||||| ||||||| |||||||
374 vaaaaaggaggggg1gggag----agrgggag----aaaaaaggaggggg1gggag 424

QY 459 SGRGLGCGAGAAAAAGAGAGGCGGCGGCGT-----SGPGGYGP- 504
: ||||||| ||||||| ||||||| ||| | ||||||| ||||||| ||||||| |||||||
Db 425 agrggg1gggag---aaaaaaggaggggggg1gggagggagaaagaggggg1 481
QY 505 GQQTSGRGLGCGAGAGAAAAAGAGAGGCGGCGTSGRGLGCGAG 558
| | : || ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 482 gsgagag---ggggagaaaaavaggggg1gggaggggg1gggag--srggg1gggag 537
QY 559 AAAA-----AAAAAGAGAGGCGGCGGCGTSGRGLGCGAGAA 598
||| ||||||| ||||||| ||| | : ||||||| ||||||| ||||||| |||||||
Db 538 aaaaaaggaggg1gggagggaggaagaaagggvrrgggg1gggag----agrgggag-- 591
QY 599 AAAAAAGAGAGGCGGCGTSGRGLGCGAGAAAAAGAGAGGCGGCGTSG 658
||||| ||||||| ||||||| ||| | : ||||||| ||||||| ||||||| |||||||
Db 592 -aaaaaaggaggggg1gggag--grrgg1gggag-----aaaaaaggaggggg1gggag 642
QY 659 GTSPPGYGPQQTSGIRIRAPSTS 683
: : | : | |
Db 643 asaa-----saasrlspqas 659

RESULT 5

AA99053 ID AAR99053 standard; Protein; 606 AA.

AA99053; AC AAR99053;

17-JAN-1997 (first entry) DT

Spider dragline variant, DP-1A.9 polymer. DE

Spider: dragline protein; variant; monomer; polymer; KM

fibre forming region; Spidroin 1; Nephila clavipes; DP1; minic; KM

DP-1A analogue; fibre; high tensile strength; elasticity; clothing; KM

rope; surgical suture; implant; reinforcement; film; coating. KM

Synthetic. OS

WO9429450-A2. PN

22-DEC-1994. PD

15-JUN-1994; 94WO-US06689. PF

15-JUN-1993; 93US-0077600. PR

(DUPO) DU PONT DE NEMOURS & CO E. I. PA

Fahnestock SR. XX

WPI: 1995-036479/05. DR

New synthetic variants of spider dragline protein - for making PT

fibres useful as clothing, surgical silk, plastic reinforcement PT

etc., also related DNA, vectors and transformed cells XX

Disclosure: Page 85-87; 168pp; English. PS

This sequence represents a synthetic spider dragline variant polymer, XX

DP-1A.9. The sequence of the DP-1A.9 monomer is given in AAR99052. CC

The polypeptide monomer is a variant based on a consensus sequence CC

derived from the fibre forming regions of spider dragline protein, CC

esp. the natural protein 1 (Spidroin 1) from Nephila clavipes. CC

DNA sequence encoding the monomer may be used in the recombinant CC

production of the variant protein in a recombinant host, e.g. E. coli CC

or Bacillus subtilis. Synthetic analogues of DP1 were designed to mimic CC

of variation among individual repeats. DP-1A analogues are composed CC

of a tandemly repeated 101 amino acid monomer which comprises four CC

repeats which differ from the consensus sequence given in AA06201, CC

according to the pattern (1)-(5) given below. This 101 amino acid CC

monomer is repeated 1-16 times in a series of analogue proteins. The CC

individual repeats differ from the consensus according to the pattern: CC

PN FR2774588-A1.
 XX 13-AUG-1999.
 XX 11-FEB-1998; 98FR-0001614.
 XX 11-FEB-1998; 98FR-0001614.
 PR (OREA) L'OREAL SA.
 PA Philippe M, Garson JC, Arraudeau JP;
 PI WPI; 1999-510729/43.
 DR
 XX Cosmetic or dermatological composition containing spider silk protein,
 PT for hair or skin care, in make-up or sunscreens
 PS Claim 7; Fig 3; 32pp; French.
 CC The present sequence represents the natural spider silk protein
 CC spiderline minor 1. The protein improves the moisturizing/softening
 CC action of the compositions. The protein, and its fragments are used
 CC in cosmetic or dermatological compositions. These compositions have
 CC use as hair or skin care products; and make-up or sunscreens.
 CC As the protein is a good, persistent film-formers on the skin
 CC of low surface density, it can be used for delivery of active
 CC agents that are generally difficult to administer, e.g. vitamins,
 CC hormones, moisturizers or agents for treating disorders of the
 CC skin and hair.
 CC
 SO Sequence 615 AA:

Query Match 39.5%; Score 1400; DB 20; Length 615;
 Best Local Similarity 50.5%; Pred. No. 5.5e-93;
 Matches 349; Conservative 18; Mismatches 204; Indels 120; Gaps 26;

QY 11 GRGMSGRGRLGSGQA-GAATAAAAGAGGCGTGGTSGRGLGCGAGAAA 69
 DB 4 gagggygrgaggygagagagagagagagaggy-----grgagaga 51
 QY 70 AAAAAAGAGGCGTGGTSGRGLGCGAGAAAAGAGAGAGAGAGAGAGAGAG 127
 DB 52 gaaaga 104
 QY 128 QGTSGRGLGSGQ-----GAGAAAAAGAGAGGCGTGGTSGRGLGCGAGAAA 182
 DB 105 aag 162
 QY 183 RGLGCGAG 241
 DB 163 Yggggygrgag 222
 QY 242 GAGGCGG-----GLGSGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAG 294
 DB 223 agggaggygrgagagagagagagagagagagagagagagagagagagag 276
 QY 295 SGRGLGCGAG 354
 DB 277 -grg--ag 332
 QY 355 AGAAAAAG 404
 DB 333 rgag 390
 QY 405 GGAAGCGTGGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
 DB 391 ag 437
 QY 464 LGAG 519
 DB 438 aag 497

QY 520 GAAAAAG 579
 DB 498 gaaaga-----gagaggy-----ggggygagagagagagagagag 538
 QY 580 GSGGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 639
 DB 539 gr-----gagagagagagagagagagagagagagagagagagagag 577
 QY 640 AAAAAAG 669
 DB 578 aaga--gagrgygrgagagagagagagagagagagagagagagagag 606

RESULT 14
 AAR9059 standard; Protein; 714 AA.
 ID AAR9059 standard; Protein; 714 AA.
 AC AAR9059;
 XX 17-JAN-1997 (first entry)
 DT Spider dragline variant, DP-2A polymer.
 DE Spider dragline variant, DP-2A polymer.
 XX Spider: dragline protein; variant; monomer; polymer;
 KW fibre forming region; Spiderline 1; Nepheila clavipes; DPL; mimic;
 KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KW rope; surgical suture; implant; reinforcement; film; coating.
 XX Synthetic.
 OS Synthetic.
 XX W09429450-A2.
 PN 22-DEC-1994.
 PD 15-JUN-1994; 94WO-US06689.
 PF 15-JUN-1993; 93US-0077600.
 PR (DUPD) DU PONT DE NEMOURS & CO E I.
 PA Fahnstock SR;
 PI WPI; 1995-036479/05.

DR New synthetic variants of spider dragline protein - for making
 XX fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc., also related DNA, vectors and transformed cells
 PS Disclosure; Page 103-105; 168pp; English.

CC This sequence represents a synthetic spider dragline variant polymer,
 CC DP-2A. The sequence of the DP-2A monomer is given in AAR9058. DP-2
 CC has a repetitive sequence and is rich in Gly and Ala. It is unlike
 CC DP-1 except for the region of consecutive Ala residues. The repetitive
 CC nature of the protein and the pattern of variation lead to the
 CC production of the consensus sequence given in AAR9085. The individual
 CC repeats differ from the consensus sequence according to a pattern which
 CC can be generalised as follows:

(1) the poly-alanine sequence varies in length from 6-10

CC residues;
 CC (2) aside from the poly-alanine sequence, individual repeats
 CC differ from the consensus repeat sequence by deletions of integral
 CC multiples of five consecutive amino acids consisting of the pentapeptide
 CC sequences GPGGY or GPGGQ. Synthetic analogues of DP2 were designed to
 CC mimic the repeating consensus sequence of the natural protein and the
 CC pattern of variation among individual repeats. The analogue DP-2A is
 CC composed of tandemly repeated 119-amino acid peptide monomers which
 CC comprises three repeats which differ according to (1)-(2) above. This
 CC 119 amino acid monomer is repeated 1-16 times in the DP2 analogues.
 CC The proteins may be used to produce fibres of high tensile strength and
 CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
 CC for implants, plastic reinforcements, films, coatings, etc.

SQ Sequence 714 AA:

Query Match 38.0%; Score 1346.5; DB 16; Length 714;
 Best Local Similarity 50.0%; Pred. No. 4e-89;
 Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

2 ASMTGGQMGKRGSMASGRGLGQGAG----AAAAAAAAGAGCGGYG-GIGSGGTSG 56
 | : | | | | | : | | | | | : | | | | | | | | | |
 Db 10 aaagppgqgppgypgypgqgppgypgqgppgsaaaaaaagp9y9p9q9p9p9y9 69
 QY 57 RGLGSGGAGAAAAAAG----GAGCGYGLGSGGTSGRGLG-GGAGAAAAA 111
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 70 p9q9p9p9saaaaaaagp9y9p9q9p9p9y9-p9q9p9p9y9p9q9p9p9saa 128
 QY 112 AAAGAGCGGYGLGSGGTSGRGLG-GGAGAAAAAAGAGCGGYGLGSGGTSG 170
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 129 aaagppgqgppgypgypgqgppgypgqgppgsaaaaaaagp9y9----p9q9 183
 QY 171 PGYGPQQTSGRGLGCGAGAAAAAAG----GAGCGYGLGSGGTSGRGLG- 225
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 184 p9y9p9q9-----p9p9saaaaaaagp9y9p9q9p9p9y9-p9q9p9p9y9 235
 QY 226 GCGAGAAAAAAGAGCGGYGLGSGGTSGRGLG-GGAGAAAAAAGAGCGG 284
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 236 gq9p9p9saaaaaaagp9q9p9p9y9-p9q9p9p9y9p9q9p9p9saaaaaaagp9 294
 QY 285 GYGGLGSGGTSGRGLGCGAGAAAAAAGAGAGCGGYGLGSGGTSGPGYGPQQT 344
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 295 y9p9-p9q9p9p9y9-p9q9p9p9saaaaaaagp9y9----p9q9p9p9y9p9q9 348
 QY 345 SGRGLGCGAGAAAAAAGAGAGCGGYGLGSGGTSGRGLG-GGAGAAAAA 403
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 349 p9y9p9p9q9p9p9saaaaaa-gp9q9p9p9y9-p9q9p9p9y9p9q9p9p9saaaaa 406
 QY 404 AGGAGCGGYGLGSGGTSGRGLGCGAGAAAAAAGAGCGGYGLGSGGTSGRG- 462
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 407 aaagpp9y9p-9q9p9p9y9p-9q9p9p9p9saaaaaaagp9y9p-9q9p9p9y9p 463
 QY 463 ---GLGCGAG-----AAAAAAAAGAGCGGYG-----GLSGGTSG----- 497
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 464 gq9p9p9y9p9q9p9p9saaaaaaagp9q9p9p9y9p9q9p9p9y9p9q9p9p9saa 523
 QY 498 -----GPGYGPQQTSGRGLGCGAGAAAAAAGAGAGCGGYGLGSGGTSGRG- 550
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 524 aaaaaagp9y9p9q9p9p9y9p9q9p9p9p9saaaaaaagp9y9p-9q9p9p9y9p 582
 QY 551 ---GLGCGAG-----AAAAAAAAGAGAGCGGYGLGSGGTSGRGLG-GGAGAAAAA 602
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 583 gq9p9p9y9p9q9p9p9p9saaaaaaagp9q9p9p9y9-p9q9p9p9y9p9q9p9p9saa 641
 QY 603 AAAGAGAGCGGYG-GIGSGGTSGRGLGCGAGAAAAAAGAGAGCGGYGLGSGGTSG 657
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 642 aaaaaagp9y9p9q9p9p9y9p9q9p9p9p9saaaaaaagp9y9p9q9p9p9y9- 700
 QY 658 QGTSGGTGPGQ 671
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 701 p9q9p9p9y9p9q9 714

RESULT 15

ID AAY40103 standard: protein: 714 AA.

XX AAY40103;

XX 19-NOV-1999 (first entry)

XX Polymer of an analogue of spider silk protein spidroine major 2.

XX Spider silk protein: spidroine major 1; cosmetic: make-up;

KW dermatological compositions: hair care; skin care; sunscreen;
 hormone; moisturizer; skin disorder; skin disorder.

XX Synthetic.
 OS Nephila clavipes.
 XX
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..119
 FT /note= "monomer unit"

FR2774588-A1.

13-AUG-1999.

11-FEB-1998; 98FR-0001614.

11-FEB-1998; 98FR-0001614.

(OREA) L'OREAL SA.

Philippe M. Garson JC, Arraudeau JP;

WPI, 1999-510729/43.

Cosmetic or dermatological composition containing spider silk protein,
 for hair or skin care, in make-up or sunscreens

Claim 8; Fig 7B; 32pp; French.

The present sequence represents a polymer of an analogue of the spider
 silk protein spidroine major 2. The protein improves the moisturizing/
 softening action of the compositions. The protein, and its fragments
 are used in cosmetic or dermatological compositions. These compositions
 have use as hair or skin care products; and make-up or sunscreens.
 As the protein is a good, persistent film-formers on the skin
 of low surface density, it can be used for delivery of active
 agents that are generally difficult to administer, e.g. vitamins,
 CC hormones, moisturizers or agents for treating disorders of the
 skin and hair.

SQ Sequence 714 AA:

Query Match 38.0%; Score 1346.5; DB 20; Length 714;
 Best Local Similarity 50.0%; Pred. No. 4e-89;
 Matches 367; Conservative 26; Mismatches 248; Indels 93; Gaps 35;

2 ASMTGGQMGKRGSMASGRGLGQGAG----AAAAAAAAGAGCGGYG-GIGSGGTSG 56
 | : | | | | | : | | | | | : | | | | | | | | | |
 Db 10 aaagppgqgppgypgypgqgppgypgqgppgsaaaaaaagp9y9p9q9p9p9y9 69
 QY 57 RGLGSGGAGAAAAAAG----GAGCGYGLGSGGTSGRGLG-GGAGAAAAA 111
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 70 p9q9p9p9saaaaaaagp9y9p9q9p9p9y9-p9q9p9p9y9p9q9p9p9saa 128
 QY 112 AAAGAGCGGYGLGSGGTSGRGLG-GGAGAAAAAAGAGAGCGGYGLGSGGTSG 170
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 129 aaagppgqgppgypgypgqgppgypgqgppgsaaaaaaagp9y9----p9q9 183
 QY 171 PGYGPQQTSGRGLGCGAGAAAAAAG----GAGCGYGLGSGGTSGRGLG- 225
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 184 p9y9p9q9-----p9p9saaaaaaagp9y9p9q9p9p9y9-p9q9p9p9y9 235
 QY 226 GCGAGAAAAAAGAGAGCGGYGLGSGGTSGRGLG-GGAGAAAAAAGAGAGG 284
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 236 gq9p9p9saaaaaaagp9q9p9p9y9-p9q9p9p9y9p9q9p9p9saaaaaaagp9 294
 QY 285 GYGGLGSGGTSGRGLGCGAGAAAAAAGAGAGCGGYGLGSGGTSGPGYGPQQT 344
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 295 y9p9-p9q9p9p9y9-p9q9p9p9saaaaaaagp9y9----p9q9p9p9y9p9q9 348
 QY 345 SGRGLGCGAGAAAAAAGAGAGCGGYGLGSGGTSGRGLG-GGAGAAAAA 403
 | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 349 p9y9p9p9q9p9p9saaaaaa-gp9q9p9p9y9-p9q9p9p9y9p9q9p9p9saaaaa 406

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:34 ; Search time 48.42 Seconds

(Without alignments)
287.483 Million cell updates/sec

Title: US-09-490-291-6

Perfect score: 3543
Sequence: 1 MASTGCGQMGKRGSMASRG.....TSGIRAPSTSFHHHHH 691Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgnt2_6/prodata/2/1aa/5A_COMB.pep:.*
2: /cgnt2_6/prodata/2/1aa/5B_COMB.pep:.*
3: /cgnt2_6/prodata/2/1aa/5A_COMB.pep:.*
4: /cgnt2_6/prodata/2/1aa/5B_COMB.pep:.*
5: /cgnt2_6/prodata/2/1aa/PCBUS_COMB.pep:.*
6: /cgnt2_6/prodata/2/1aa/backfile1.pep:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	60.3	718	1	US-08-425-069-2
2	2138	60.3	718	2	US-08-317-844B-2
3	2138	60.3	747	4	US-09-034-177-3
4	1405	39.7	832	1	US-08-209-747-2
5	1405	39.7	832	1	US-08-458-298-2
6	1241.5	35.0	1177	1	US-07-609-716-31
7	1241.5	35.0	1177	1	US-08-175-155-29
8	1241.5	35.0	1177	1	US-08-477-509B-64
9	1241.5	35.0	1177	2	US-08-707-237A-35
10	1241.5	35.0	1177	3	US-08-482-085B-64
11	1241.5	35.0	1177	4	US-08-475-411A-31
12	1241.5	35.0	1177	4	US-08-478-029A-31
13	1184.5	33.4	1059	2	US-08-175-155-48
14	1184.5	33.4	1059	2	US-08-707-237A-54
15	1184.5	33.4	1101	1	US-08-477-509B-83
16	1184.5	33.4	1101	3	US-08-482-085B-83
17	1157	32.7	738	3	US-08-864-038A-3
18	1138.5	32.1	595	2	US-08-425-069-4
19	1138.5	32.1	595	2	US-08-317-844B-4
20	1138.5	32.1	1038	1	US-07-609-716-36
21	1138.5	32.1	1038	4	US-08-475-411A-36
22	1138.5	32.1	1038	4	US-08-478-029A-36
23	1137.5	32.1	1332	4	US-07-609-716-41
24	1137.5	32.1	1332	4	US-08-475-411A-41
25	1137.5	32.1	1332	4	US-08-478-029A-41
26	1135	32.0	979	1	US-08-477-509B-89
27	1135	32.0	979	3	US-08-482-085B-89

28	1135	32.0	1050	1	US-08-175-155-54	Sequence 54, App1
29	1114	31.4	1018	1	US-08-089-862-11	Sequence 11, App1
30	1114	31.4	1018	1	US-08-587-333-18	Sequence 18, App1
31	1114	31.4	1018	5	PCT-US94-07776-16	Sequence 16, App1
32	1104	31.2	2100	3	US-08-477-509B-80	Sequence 80, App1
33	1104	31.2	2100	3	US-08-482-085B-80	Sequence 80, App1
34	1104	31.2	2107	1	US-08-175-155-45	Sequence 45, App1
35	1104	31.2	2107	2	US-08-707-237A-51	Sequence 51, App1
36	1103	31.1	907	2	US-09-010-928B-4	Sequence 4, App1
37	1099	31.0	945	1	US-08-089-862-6	Sequence 6, App1
38	1099	31.0	945	1	US-08-587-333-13	Sequence 13, App1
39	1099	31.0	945	5	PCT-US94-07776-11	Sequence 11, App1
40	1090.5	30.8	766	1	US-08-175-155-53	Sequence 53, App1
41	1090.5	30.8	766	1	US-08-477-509B-88	Sequence 88, App1
42	1090.5	30.8	766	2	US-08-707-237A-61	Sequence 61, App1
43	1090.5	30.8	766	3	US-08-482-085B-88	Sequence 88, App1
44	1084	30.6	745	2	US-09-010-928B-28	Sequence 28, App1
45	1081.5	30.5	870	2	US-09-010-928B-2	Sequence 2, App1

ALIGNMENTS

RESULT 1
US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himan, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match 60.3%; Score 2138; DB 1; Length 718;
Best Local Similarity 66.3%; Pred. No. 3.5e-147;
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;
0Y 25 OGAGAAAAAAGAGAGCGTGGTGGT--SGRGIGGAG-AAAAAAAAAGAGAG 81

TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 832 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: ORGANISM: N. clavipes
: TISSUE TYPE: minor amputate gland
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..309
: US-08-209-747-2

Query Match 39.7%; Score 1405; DB 1; Length 832;
Best Local Similarity 50.1%; Pred. No. 2,1e-94;
Matches 349; Conservative 17; Mismatches 207; Indels 124; Gaps 26;

QY 23 GGCGAGAAAAAGAGAGGCGTGGCGTSGRGLGCGGAGAAAAAGAGAGG 82
DB 79 GGGYAG---AGAVAAAAAGAGAGY---RGAGYGGGCGYGAAGAGAAAA--GAGAG 129
QY 83 GYGGLSGGTSGRGLGCGGAGAAAAAGAGAGGCGY---GLGSGTSGRGLG 137
DB 130 GAGY-----GRGAGAGAGAAAAAGAGAGAGYGGGCGYGAAGAGAAAAAGAGAG 183
QY 138 GGGAGAAAAAGAGAGGCGYGGGCGTSGPGYGPQGTSGRGLGCGGAGAAAA 197
DB 184 GYGRAGAGAGAAA--GAGAGGCGGCGYGAAGAGAAAAAGAGSGAGGAGAGAG 241
QY 198 AAAAAAGAGGCGYGGGCGTSGRGLGCGGAGAAAA-----AAAAG 241
DB 242 GAAAGAGAGAGY-----GGGCGAGAGAGAAAAAAXXXXXXXXXXXXXGAGAG 293
QY 242 GAGGCGY---GLGSGTSGRGLGCGGAGAAAAAGAGAGGCGY--- 288
DB 294 YGGGCGYGAAGAGAGAGAGAGAGAGGCGRAGAGAGAGAGAGAGAGAGAGAG 353
QY 289 ---LGSQGTSGRGLG---GGAGAGAAAAAGAGAGGCGYGGGCGTSGPGYGP 341
DB 354 GAAAAAG 412
QY 342 QGTSGRGLGCGGAGAAAAAGAGAGGCGYGGGCGTSGRGLG 390
DB 413 AAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 470
QY 391 OGAGAAAAAGAGAGGCGYGGGCGTSGRGLGCGGAGAAAAAGAGAGGCGY 450
DB 471 YRGAGAGAGAGAGAGAGAGG-----AGYCGGCGYGAAGAGAGAGAGAGAG 521
QY 451 G-GLSGGTSGRGLGCGGAGAGAAAAAGAGAGGCGY---LGSQGTSGPG-YGPG 505
DB 522 GAGAG---AG 577
QY 506 QGTSGRGLGCGGAGAAAAAGAGAGGCGY---GLG-----QGTSGRGLG-GG 556
DB 578 YSRGAG 637
QY 557 AGAGAAAAAGAGAGAGGCGYGGGCGTSGRGLGCGGAGAAAAAGAGAGGCGY 616
DB 638 AGAGAGAGAGAGAGAGAGY-----GGGCGYGAAGAGAGAGAGAGAGAGAG 683
QY 617 LGSQGTSGRGLGCGGAGAGAAAAAGAGAGGCGY 653
DB 684 -RGAGAGYGGGCGYGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717

,RESULT 5

US-08-458-298-2
: Sequence 2, Application US/08458298
: Patent No. 575677
: GENERAL INFORMATION:
: APPLICANT: Lewis, Randolph V.
: TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider
: TITLE OF INVENTION: Silk Proteins
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,298
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,747
: FILING DATE: 14-MAR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1447-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 832 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: ORGANISM: N. clavipes
: TISSUE TYPE: minor amputate gland
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..309
: US-08-458-298-2

Query Match 39.7%; Score 1405; DB 1; Length 832;
Best Local Similarity 50.1%; Pred. No. 2,1e-94;
Matches 349; Conservative 17; Mismatches 207; Indels 124; Gaps 26;

QY 23 GGCGAGAAAAAGAGAGGCGYGGGCGTSGRGLGCGGAGAAAAAGAGAGG 82
DB 79 GGGYAG---AGAVAAAAAGAGAGY---RGAGYGGGCGYGAAGAGAAAA--GAGAG 129
QY 83 GYGGLSGGTSGRGLGCGGAGAAAAAGAGAGGCGY---GLGSGTSGRGLG 137
DB 130 GAGY-----GRGAGAGAGAAAAAGAGAGAGYGGGCGYGAAGAGAAAAAGAGAG 183
QY 138 GGGAGAAAAAGAGAGGCGYGGGCGTSGPGYGPQGTSGRGLGCGGAGAAAA 197
DB 184 GYGRAGAGAGAAA--GAGAGGCGGCGYGAAGAGAAAAAGAGSGAGGAGAGAG 241
QY 198 AAAAAAGAGGCGYGGGCGTSGRGLGCGGAGAAAA-----AAAAG 241
DB 242 GAAAGAGAGAGY-----GGGCGAGAGAGAAAAAAXXXXXXXXXXXXXGAGAG 293


```

: FILING DATE: 09-NOV-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/114,618
: FILING DATE: 29-OCT-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 06/927,258
: FILING DATE: 04-NOV-1986
: ATTORNEY/AGENT INFORMATION:
: NAME: Treacartin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-781-1989
: TELEFAX: 415-398-3249
: INFORMATION FOR SEQ ID NO: 31:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1177 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-478-029A-31

```

```

Query Match      35.0%; Score 1241.5; DB 4; Length 1177;
Best Local Similarity 40.3%; Pred. No. 1.8e-82;
Matches 292; Conservative 58; Mismatches 313; Indels 61; Gaps 12;

```

```

OY 7 GQOMGRGSMASGRGLGSGGAGAAAAAAGAGAGGCGGCGTSGRGLGSGGAG 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 GSGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 67 AAAAAAAAAA-----GAGGCGTGGCGTSGRGLGSGGAGAAAAA 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 SGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 114 AGGAGAG--GTC-----GLCSQGTSGRGLGCGGAGAAAAAAGAGGCGG 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 GAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 165 SGTGSGPGGCGGCGTSGRGLGCGGAGAAAAAAGAGAGGCGGCGTSGR 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 AGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 222 G---GLGSGGAGAAAAAAGAGAGGCGGCGTSGRGLGCGGAGAAAAA 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 279 GAGAGGCGT-----GLCSQGTSGRGLGCGGAGAAAAAAGAGGCGG 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 AGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 332 TSGPGGCGGCGTSGRGLGCGGAGAAAAAAGAGAGGCGGCGTSGR 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 GSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 387 -GLGSGGAGAAAAAAGAGAGGCGGCGTSGRGLGCGGAGAAAAAAG 445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 SGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 446 GCGGCGT-----GLCSQGTSGRGLGCGGAGAAAAAAGAGGCGG 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 499 PGCGTSGCGTSGRGLGCGGAGAAAAAAGAGAGGCGGCGTSGR 552
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 664
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 553 GCGGAGAAAAAAGAGAGGCGGCGTSGRGLGCGGAGAAAAAAGAGAG 612
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 724
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 613 GTC-----GLCSQGTSGRGLGCGGAGAAAAAAGAGGCGGCG 665
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 725 GYGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 784
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 666 YGPG 669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 YGAG 788
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-08-175-155-48
: Sequence 48, Application US/08175155
: Patent No. 5641648
: GENERAL INFORMATION:
: APPLICANT: Ferrari, Franco A.
: APPLICANT: Cappello, Joseph
: APPLICANT: Crissman, John W.
: APPLICANT: Dorman, Mary A.
: TITLE OF INVENTION: Methods for Preparing Synthetic
: NUMBER OF SEQUENCES: 69
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohbach, Teet, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: US
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/175.155
: FILING DATE: 29-DEC-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Rowland, Bertram I.
: REGISTRATION NUMBER: 20015
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-781-1989
: TELEFAX: 415-398-3249
: INFORMATION FOR SEQ ID NO: 48:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1059 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-175-155-48

```

```

Query Match      33.4%; Score 1184.5; DB 1; Length 1059;
Best Local Similarity 40.7%; Pred. No. 1.8e-78;
Matches 272; Conservative 60; Mismatches 323; Indels 13; Gaps 7;

```

```

OY 7 GQOMGRGSMASGRGLG--GCGAGAAAAAAGAGAGGCGGCGTSGRGLGSG 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 GSGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 66 GAAAAAAGAGAGAGGCGGCGTSGRGLGCGGAGAAAAAAGAGAGGCG 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 GSGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 126 GSGTSGRGLGCGGAGAAAAAAGAGAGGCGGCGTSGRGLGSGGAG 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 GSGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 185 GLGCGAGAAAAAAGAGAGGCGGCGTSGRGLGCGGAGAAAAAAGAGAG 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 AGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 620
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

01 245 OGGVGGGCGSOGTSGRGLGCGGAGAAAAAAGAGAGGCGVGGGCGSOGTSGRGLGCGG 304
02 245 OGGVGGGCGSOGTSGRGLGCGGAGAAAAAAGAGAGGCGVGGGCGSOGTSGRGLGCGG 304
03 621 AGAGSGAGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 680
04 305 AGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 363
05 601 AGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 737
06 364 AAAGGAGGCGVGGGCGSOGTSGRGLGCGGAGAAAAAAGAGAGGCGVGGGCGSOGTSGR 423
07 738 GSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 797
08 424 GGLGGGAGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 483
09 798 GSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 857
10 484 GGGGCGGCGSOGTSGRPG - GYGPGCGGTSGRGLGCGGAGAAAAAAGAGAGAGAGAGAG 542
11 858 GSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 914
12 543 SGTGSRGGLGCGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
13 915 AGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 974
14 603 AAAAGAGAGCGGCGGCGSOGTSGRGLGCGGAGAAAAAAGAGAGAGAGAGAGAGAGAG 662
15 975 AGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1034
16 663 PGGCGPGC 670
17 1035 AGAMPGR 1042
18
19 RESULT 14
20 US-08-707-237A-54
21 Sequence 54, Application US/08707237A
22 Patent No. 5830713
23 GENERAL INFORMATION:
24 APPLICANT: Ferrari, Franco A.
25 APPLICANT: Capello, Joseph
26 APPLICANT: Criseman, John W.
27 APPLICANT: Dorman, Mary A.
28 TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
29 TITLE OF INVENTION: REPETITIVE DNA
30 NUMBER OF SEQUENCES: 108
31 CORRESPONDENCE ADDRESS:
32 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
33 STREET: Four Embarcadero Center, Suite 3400
34 CITY: San Francisco
35 STATE: California
36 COUNTRY: United States
37 ZIP: 94111-4187
38 COMPUTER READABLE FORM:
39 MEDIUM TYPE: Floppy disk
40 OPERATING SYSTEM: IBM PC compatible
41 SOFTWARE: PatentIn Release #1.0, Version #1.30
42 CURRENT APPLICATION DATA:
43 APPLICATION NUMBER: US/08/707,237A
44 FILING DATE: 03-SEP-1996
45 CLASSIFICATION: 435
46 PRIOR APPLICATION DATA:
47 APPLICATION NUMBER: US 08/175,155
48 FILING DATE: 29-DEC-1993
49 PRIOR APPLICATION DATA:
50 APPLICATION NUMBER: US 08/053,049
51 FILING DATE: 22-APR-1993
52 PRIOR APPLICATION DATA:
53 APPLICATION NUMBER: US 07/609,716
54 FILING DATE: 06-NOV-1990
55 PRIOR APPLICATION DATA:
56 APPLICATION NUMBER: US 07/269,429
57 FILING DATE: 09-NOV-1988

```

[illegible]

This Page Blank (uspto)

Result No.	Score	Query Match	length	DB	ID	Description
1	2138	60.3	718	2	A36068	major amputate fib
2	1388	39.1	2639	2	T31328	fibroin - Chinese
3	1281.5	36.2	1901	2	D70806	hypothetical glyci
4	1260.5	35.6	1489	2	D70807	hypothetical glyci
5	1242	35.1	1079	2	B70807	hypothetical glyci
6	1198.5	33.8	1306	2	A70934	hypothetical glyci
7	1196.5	33.8	1381	2	E70869	hypothetical glyci
8	1180	33.3	1660	2	A70869	hypothetical glyci
9	1158.5	32.7	1538	2	E70846	hypothetical glyci
10	1152	32.5	1329	2	E70917	hypothetical glyci
11	1151	32.5	863	2	A70896	hypothetical glyci
12	1138.5	32.1	637	2	A44112	spidroin 2, dregli
13	1114.5	31.5	749	2	A70812	hypothetical glyci
14	1110.5	31.3	914	2	H70987	hypothetical glyci
15	1083	30.6	882	2	B70812	hypothetical glyci
16	1049.5	29.6	778	2	F70963	hypothetical glyci
17	1038.5	29.3	767	2	E70895	hypothetical glyci
18	1032.5	29.1	837	2	E70835	hypothetical glyci
19	1032	29.1	957	2	D70835	hypothetical glyci
20	1014	28.6	783	2	E70824	hypothetical glyci
21	1007	28.4	731	2	C70974	hypothetical glyci
22	992	28.0	714	2	A70807	hypothetical glyci
23	989	27.9	741	2	G70917	hypothetical glyci
24	988.5	27.9	801	2	F70824	hypothetical glyci
25	977.5	27.6	1011	2	F70620	hypothetical glyci
26	976	27.5	860	1	EAMS	elastin precursor
27	939	26.5	864	1	EAPR	elastin precursor
28	937	26.4	923	2	E70820	hypothetical glyci
29	923.5	26.1	667	2	A70893	hypothetical glyci

30	903.5	25.5	694	2	G70868	hypothetical glyci
31	893	25.2	484	2	G70846	hypothetical glyci
32	869	24.5	603	2	A70770	hypothetical glyci
33	853.5	24.1	465	1	S01820	hypothetical glyci
34	848.5	23.9	1585	1	T31611	glycine-rich cell
35	833.5	23.5	576	2	A70900	hypothetical glyci
36	817.5	23.1	584	2	G70804	hypothetical glyci
37	809	22.8	770	2	S59623	tropoelastin - she
38	808.5	22.7	618	2	A70989	hypothetical glyci
39	804	22.7	639	2	D70931	hypothetical glyci
40	801	22.6	606	2	H70816	hypothetical glyci
41	800	22.6	591	2	B70523	hypothetical glyci
42	786.5	22.2	747	1	EABO	elastin precursor
43	777.5	21.9	588	2	F705971	hypothetical glyci
44	776.5	21.9	615	2	H70589	hypothetical glyci
45	763	21.5	641	1	Q0BE31	nuclear antigen EB

ALIGNMENTS

RESULT 1
A36068
major amputulate fibroin protein - orb spider (Nephila clavipes) (fragment)
C:Species: Nephila clavipes
C:Date: 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Sep-1997
C:Accession: A36068
R:Xu, M.; Lewis, R.V.
Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990
A:Title: Structure of a protein superfiber: spider dragline silk.
A:Reference number: A36068; MUID:90384959
A:Accession: A36068
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-718 <XU>
A:Cross-references: GB:M37137; NID:g159711; PID:g159712
!Note: the authors translated the codon GGT for residue 252 as Gln, GTA for residue

Query Match	60.3%;	Score 2138;	DB 2;	Length 718;
-------------	--------	-------------	-------	-------------

Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;

[illegible]

Db	371	AG--AAVAAAAAGGAGGCGYGGLGSQG----	AGRCGCGAG----	AAAAAAGAGAGRRGCG	419
Qy	453	LGSGCTSGRGLGGGOGAGAAAAAAGAGGCGGCGGLGSQGT-----	SGP	499	
Db	420	LGNGG-AGRGGLGSGOGAG--AAAAAAGAGGCGGCGGLGNGGAGRGCGGCGAAAAAGCAGQ		475	
Qy	500	GGYGP-GQOTSGRGGLGSGGAGAAAAA-----	AGCAGGCGYGGLGSQSTSGRGL	552	
Db	476	GGYGGLGSGOGAGR--GGCGAGAAAAAAGCAGCGCIRCGAGGCGYGGLGSQG--SGRGL		531	
Qy	553	GGCGAGAAAA-----	AAAAAGAGGCGYGGLGSQSTSGRGLG	592	
Db	532	GGCGAGAAAAAGGAGCGGGLGGCGAGCGAGAAAAAGCYRGCGTGGLSQG----	AGRCG	587	
Qy	593	QGAGAAAAAAGAGAGCGGCGYGGLGSQSTSGRGLGSGGAGAAAAAAGGAGCGG		652	
Db	588	QGAG-----AAAAAAGAGAGCGCGYGGLGCGV--GRCGLGCGAG-----	AAAAAGCGCGY	636	
Qy	653	GGLSQGTSGPCCGYPGQGTGIRIRAPSTS	683		
Db	637	GGVGAGASAA-----	SAASTRLLSPQAS	659	

RESULT 2
T31328
fibroin - Chinese oak silkwmoth
C:Species: Antherea pernyi (Chinese oak silkwmoth)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31328
R:Sezutsu, H.; Tamura, T.; Yukuhito, K.
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, Antherea pernyi
A:Reference number: 220995
A:Accession: T31328
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2639 <SE2>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AC32606.1
C:Genetics:
A:Introns: 14/3

Query Match	39.1%	Score 1386;	DB 2;	Length 2639;
Best Local Similarity	40.4%	Pred. No. 6.4e-71;		
Matches 412; Conservative	21;	Mismatches 236;	Indels 352;	Gaps 40;

```

Oy 2 ASMTGGCGMGSGSMASGGGGLGCGGAGAGAAAAAAAAAAGAGCGGCGT----- 48
    || || || || || || || || || || || || || || || || || || || || ||
Db 585 ASGAGGGGGGTG---WGCGTGTSDSAAAAAAAAAAGCGAGGAGGGTGMCKDGTGSDS 641
Oy 49 -----LGSGTGTSGRGLGCGGAGAAAAAAAAAAGCAGCGGTGTGSGT-S 94
    || || || || || || || || || || || || || || || || || || || || ||
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 642 AAAAAAAAAAASGARSGGTGTGTGSDSAAAAAAAAAAGSAGGTGTG---GTGAG 697
Oy 95 RGLGCGGAGAAAAAAAA---AGCAGCGGTGTGCGT-----SGRGLG 138
    || || || || || || || || || || || || || || || || || || || || ||
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 698 DCGYSDSAAAAAAAAAAGSGAGRGDGTGTGSSSSAAAAAAAAASARRAGHDSAA 757
Oy 139 QGAGAAAAAAAAAAGAG--QGCY---GTLGSGT-----SGRGTGTGPGQ 179
    || || || || || || || || || || || || || || || || || || || || ||
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 758 SAAAAAAAAAAGAGSGGSGGTGTGSDSAAAAAAAAAAGSGAGGAGGTG 817
Oy 180 TSGRGLGCGGAGAAAAAAAA---GCAGCGGTGTGCGT-----SGRGL 224
    || || || || || || || || || || || || || || || || || || || || ||
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 818 W-GDGTGTSDSAAAAAAAAAAGAGSGGTGTGTSDSAAAAAAAAAAGSAGAGT 876
Oy 225 GG-----QGAGAAAAAAAA---AGCAGCGGTGTGCGT-----S 257
    || || || || || || || || || || || || || || || || || || || || ||
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 877 GCGTGTGSDSAAAAAAAAAAGSGAGRGDGTGTGSSSAAAAAAAAASARRA 936
Oy 258 RGLGCGGAGAAAAAAAAAAGAG--QGCY---GTLGSGT-----SGR 298
    || || || || || || || || || || || || || || || || || || || || ||
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 937 GHDSAAAGSAAAAAAAAAAGAGSGGTGTGSDSAAAAAAAAAAGSGAG 996

```

[illegible]

RESULT	3
F70806	
hypothetical glycine-rich protein RV3508 - Mycobacterium tuberculosis (strain H37RV)	

C:Species: mytilus collagens
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70806
R:COLE, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno-
A:Reference number: A70500; MUID:98295987
A:Accession: F70806
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1901 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:93261554; PIDN:CAAL17745.1; PID:g2922
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3508
C:Superfamily: collagen alpha 1(IV) chain

Query Match	36.28;	Score 1281.5;	DB 2;	Length 1901;
Best Local Similarity	39.58;	Pred. No. 3.6e-65;		
Matches 354; Conservative	28;	Mismatches 273;	Indels 241;	Gaps 39;

```

QY 7 GQOMGRSMASSRGGLGCG-----GAGAAAAAAGAGAGGCGGGLGCG----- 52
Db 280 GGGGAGACAGSDGAGLGGTGTGTGGAGAGRGALLLLAGCGGGLGAGCGGTGAG 339
QY 53 -----GTSRGLGCG-----GAGAAA-----AAAAAAG--AGCGYGLGCG 90
Db 340 GDEVLGAGVGTGKGGGAGAGAGAAAGQLFSAGAGAGAVGCGTGGGAGAGAA 399
QY 91 G-----TSGRGLGCGAGAAAAAAGAGAGCGGCGGGLGCGTSGR-- 133
Db 400 GADAPASTGLTGTGFGAGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
QY 134 -----GGLGCGG--AGAAAAAAG-----AGCGYGLGCGT 168
Db 460 ASGITGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
QY 169 -----SGCGYCPGCGTSGRGLGCGAGAGAAAAAAGAGAGAG-- 207
Db 520 DATATGATGTGTGFGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
QY 208 --GGYGLGCGTSGRGLGCG--AGAAAAAAG--AGCGYGLGCGT--SGR 259
Db 579 GIGGTGCGTGGK--GAGAGAGCGGSSGAGAGAGAGAGAGAGAGAGAGAGAG 636
QY 260 GGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 311
Db 637 GAGAGGTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
QY 312 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355
Db 697 GSSATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756
QY 356 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
Db 757 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 816
QY 394 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 435
Db 817 FDGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
QY 436 -----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
Db 877 NGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
QY 476 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
Db 937 SINAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 993
QY 532 GAGCGGGLGCGT-----GCGGLGCGAGAAAAAAGAGAGAGAGAGAG 580
Db 994 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1053
QY 581 SGGTSG--RGGLGCG--GAGAAA-----AAAAAAGAGAGAGAGAGAG 621
Db 1054 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
QY 622 --TSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
Db 1114 NFGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169

```

RESULT 4
 D70807
 Hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: D70807
 R:Coile, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; M01D:9825987
 A:Accession: D70807
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1489 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; M1D:93261554; P1D:CA117751.1; P1D:9292
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3514
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 35.6%; Score 1260.5; DB 2; Length 1489;
 Best Local Similarity 38.5%; Pred. No. 4,36-64;
 Matches 344; Conservative 35; Mismatches 274; Indels 241; Gaps 37;

```

QY 6 GQOMGRSMASSRGGLGCG-----GAGAAAAA-----AA 35
Db 214 GGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273
QY 36 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 90
Db 274 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
QY 91 -----GTSRGLGCG-----GAGAAAAAAGAGAGAGAGAGAGAGAGAG 131
Db 334 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391
QY 132 GCGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 175
Db 392 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 451
QY 176 PGCGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 227
Db 452 GADPPTGCGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 511
QY 228 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 255
Db 512 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 571
QY 256 --TSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
Db 572 NPTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631
QY 310 AAA-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 344
Db 632 ADAPGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 691
QY 345 SGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
Db 692 -GIGGTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 750
QY 404 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
Db 751 GGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 810
QY 460 GCGGLGCG-----GAGAAAAA-----AAAAAGAGAGAGAGAGAGAGAGAG 483
Db 811 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 870
QY 484 -----GCGYGLGCGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 523
Db 871 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 927
QY 524 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 582
Db 928 SGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 985
QY 583 GTSRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631
Db 986 SAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1044

```

```

QY      632  -----GAGGAAAA-----AAAAAGAGGCGGCGG-----SQGTSNGPGCG 667
          ||| : | | ||| || | : ||| || |
Db      1045 GAGKGAGGSSGAGGTGSCGAGAGGQGGAGAGISFSGNSNGTGTGTGGV 1098

RESULT      5

B70807
hypothetical glycine-rich protein RV3512 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
: Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Kajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: B70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1079 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:93261554; PIDN:CAA17749.1; PID:92924444
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3512
A:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

```

[illegible][illegible]

RESULT 6
A:RV0934
hypothetical glycine-rich protein RV0578c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70934
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference numbers: A70500; MUID:98295987
A:Accession: A70934
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1306 <COL>
A:Cross-references: GB:AL021942; GB:AL123456; NID:3242298; PID:CA117449.1; PID:g2930
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV0578c
C:Superfamily: collagen alpha 1(IV) chain

Query Match	33.8%	Score 1198.5	DB 2	Length 1306
Best Local Similarity	38.4%	Pred. No. 1.2e-60		
Matches 323	Conservative 37	Mismatches 293	Indels 189	Gaps 34
QY	11	GGGSMASGNGGCGAGGGGAGAGAAAAAAGG----	AGGGGAGGLGSGGTSGRGLGCGGAG	66
DB	399	GTTSTSGGGGGGKGGKADATSSGQTGANGGRGGDGGGYGNGGAG--	GAGRGKAGAGLDFG	456
QY	67	AAAAAAAAAAGG--AGCGGYGG--	LGSGGTSGRGLGCGGA-----	GAAAAAAAAA 114
DB	457	SEAPRPFGAGAGTGGGAGAGNGGTQADGGTGAGAGAGBGGSGAGSITGPMASAPGAGSP		516
QY	115	GG-----AGCGGTGGL-----	GSQGTSGRGLGCGGA-----	GA AAAA--- 146
DB	517	GGNGCGNGPGGAGGREGAGAGLALAAASGQNSGAGAGDGGAGCGNGGTGPGNGHGAAAGALGV		576
QY	147	--AAAAAGGAAGGAGGGLGSGGTSG--	PGTGC-TGQ--	GTSGRGLGSGGAGAAAAAA 199
DB	577	NGGVGAGAGHGDDPGVGAGAGGGGSGSTPPGANGARGNTPTTSGGNGNGSGRGGADATGFCGT		636
QY	200	AAAGG-----	AGGGGAGGLGSG-----	GTSGRGLGAGGAGAAAA 234
DB	637	GASGGGCGDGGGLVNGMGAGAGAGNGSKGLPLGLRLCNPGLDGGTGGNGAGGSGGAMACN		696
QY	235	AAAAAAGG--AGCGGYGGLGSGQ-----	GTSGRGLGAGGQ--	AGAAAAAAA 277
DB	697	GGTGGAGAGTGGVGTGTGGSSSDGVNGSSAGADHPGGTGGVGTGGKGGGDDGAAPKCV		756
QY	278	AGGAGGGGAGGLGSGGTSGRGLGSGGAGAAAAAAGG--	GAGCGGTGGLGSGQ--	GTSCP 335
DB	757	ASGQPGGAGGDDG--	GTGTVGNGNGRGIDGADGATAGARGDDGAGAGGKGGRGRTGGR	814
QY	336	GGTGP-----	GGQTSRGLGAGG-----	AGAAAAAAA 364
DB	815	GGAGPAGTTGSGAGAGNGGSGGTGGDDPGDGGCANGASVFTNNGIGGNGGNGNAGPSGAG		874

Df 485 RCGAGCGSGGSCDCCGGAGAGW-----LFGDCAGCNGGAAAAAGAGGACGGGGNGG 539
| | | | | : | | | | |
Oy 369 -AGCGYTGGLSOGTS-----GRGCLGGCGAGAAAAAAA----AAGCAQGGYTGGLSQ 418
| | | | | : | | | | |
Db 540 NGCGNGCGNCGNATCATGMVLGNMGAGGCGATACAGCAGANGVSTTGGGTGGNGIG-- 597
| | | | | : | | | | |
Oy 419 GTSRRLGLGGGA--GAAAAAAAAAAAGAG-OCGYGIG-----SQTSGRGLGCG-QAG 470
| | | | | : | | | | |
Db 598 GTGGGGAGGANAALLGVGAGAGHAGASGAGDRGAGGTGFTSPDGAGCDGCGNNGAG 657
| | | | | : | | | | |
Oy 471 AAAAAAAAAAAGAGCGGCGYTGGLSOGTSGPGYGPGQOTSGRGLGCGAGAAAAAAA 530
| | | | | : | | | | |
Db 658 GTGGLLPAGAGNNGPBGSGGAAIDIG--GNGAGNGGCTDNGNGSGSGG---AGSGDG 712
| | | | | : | | | | |
Oy 531 GCAGCGYTGGLSOGTSGRGLGCGAGAAAAAAAAGAGCGCYTGGLSOGTSGRGL 590
| | | | | : | | | | |
Db 713 GGAGNGAMLEFNGAGAGCGCGKGNAG-----GGTGGSGFLPLNGSGDGD 762
| | | | | : | | | | |
Oy 591 GCGGAGAAAAAAAAGAGCGGCGYTGGLSOGTSGRGLGCGAGAAAAAAAAGAGAGG 650
| | | | | : | | | | |
Db 763 GGNAPGCVLYGNGAG--GGSSGGTGGPGAT--GGAGKGCGDGDQLTIGDGGNGNG 818
| | | | | : | | | | |
Oy 651 GYGLSGOGTSGPGYGPGQOTSGT 675
| | | | | : | | | | |
Db 819 GAGGTG--GTPGPGPGSGSLGL 841
| | | | | : | | | | |

RESULT 12
A44112
spidrolin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)
N:Alternate names: silk fibroin, dragline
C:Species: Nephila clavipes
C>Date: 31-Dec-1993 #sequence-revision 31-Dec-1993 #text-change 20-Mar-1998
C:Accession: A44112; S27824
R:Hilman, M.B.; Lewis, R.V.
J. Biol. Chem. 267, 19320-19324, 1992
A>Title: Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes
A:Reference number: A44112; MUID:92406876
A:Accession: A44112
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-627 <HIN>
A:Cross-references: GB:M92913; NID:g159713; PID:g159714
A>Note: Sequence extracted from NCBI backbone (NCBIP:113893)
R:Hilman, M.B.; Lewis, R.V.
submitted to the EMBL Data Library, May 1992
A:Description: Isolation of a clone encoding a second dragline silk fibroin: Nephila cl
A:Reference number: S27824
A:Accession: S27824
A:Molecule type: mRNA
A:Residues: 19-627 <HTZ>
A:Cross-references: EMBL:M92913

```

Query Match Similarity 32.1%; Score 1138.5; DB 2; Length 627;
Best Local Similarity 47.0%; Pred. No. 1.6e-57;
Matches 325; Conservative 36; Mismatches 190; Indels 141; Gaps 36;

OY 23 GGQGGAGAAAAAAGAGAGCGGCGTSGRGLGAGCGAGAAAAAAGC----G 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 GCGGPGGQ-----GPGGYGP-GQQGPGSGP-----SAAAAAAGGPGGYG 42
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 79 AGCGGCGGCG-SQGTSGRGLGCGGAGAAAAAAGAGAGCGGCGTSGRGLSQ----GTSGR 133
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 43 PGQGGPGCGTGGCGGCGRGRGPGCGGPGSPGPGSAAAAA--GSGQGGPGGCGPRQGGPGGCG 101
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 134 GGLGCGGAGAAAAAAGAGAGCGGCGTSGTSGPGCGPGCGTSGRGLG----- 187
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 102 GQGGGPGSGPSSAAAAAASAESGQGGPGGCG--PGQGGPGGCGPGGQ--GPGGCGGCGG 158
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 188 -CGAGAGAAAAAAGAGAGCGGCGTSGRGLGCGGCGTSGRGLG-----GAGCGAGAGAAAAAAG 242
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

RESULT 13
 A70812
 Hypothesized glycine-rich protein RV0833 - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_rev1ston 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70812
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 J.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:96295987
 A:Accession: A70812
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-749 <COL>
 A:Cross-references: GB:AL022004; GB:AL123456; NID:33261550; PIDN:CAAL7639.1; PID:g29201
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: RV0833
 C:Superfamily: elastin

A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAL7640.1; PID:g291689
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV0834c
C:Superfamily: elastin

```
Query Match          30.6%; Score 1083; DB 2; Length 882;
Best Local Similarity 39.7%; Pred. No. 2,7e-54;
Matches 306; Conservative 28; Mismatches 306; Indels 130; Gaps 33;

OY 7 GQMGKRGSMASGRGLGQGGAGAAAAAAGAGAGCGGCGYGLGSQGTSGRGLGSGGAG 66
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 133 GEAGGPGGMLLNGNGSGSGAPGQTGGAGAAAGLLRHGTGAG--GTASGKGGTGW 190
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 67 AAAAAAAAAAGAGCGGCGYGLGSGQ---GTSRGLG--OGAGAAAAAAAAAGAG--G 118
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 191 LMSGGAGAGAGSG--GSGGAGAGNLMFGTIGNGAGAGAAAGVGVGAGAGGALVA 249
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 119 QGGYGGIGSGGTSGRGLGSGGAGAAAAAAGAGAGCGGCGYGLGSQGTSGPGYG--- 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 250 IGGAGGAGCAATTGTGAGAGAGSN-ALGLFLGLGSGGQGGDSAMSGAGAGSGGAA 308
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 176 -PGQOTSGRGLGCGAGAGAAAAAAGAGAG-----OGGYGLGSQGTSGRG 222
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 309 SPFGIDIGGAGGHG--GAGTNGGAGAGAGSGSYFALDLSWAGAGNGAATTGTG 367
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 223 GLGCGAGAAAAAAGAGAGCGGCGYGLG--SQGTSGRGLGSGGAGAAAA---AAA 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 368 GAG--CTGGFNAHPFIFGAAAYGAGGAGGATGAGTGGTGGVAGGFALGVYVGA 425
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 276 AAAGAG----GQGGYGLGSQGTSGRGLGSGGAGAAAAAAGAG-----GAGQ 321
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 GGAGGAATETGIGAGGLGVLLGAGAGGPG--GAASAGSGHGTGCDALGLIAGI 484
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 322 GGYGGLG-----SQGTSGPGCGPQGTSGRGLGSGGAGAAAAAAGAGCGGCGY 376
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 485 GGVGGVGAATDTGGNGAGGSGTG-LLGVGAGAGHGG--ASVGTGSGGAGGDFGF 541
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 377 LGSQGTSGRGLG-----GQAGAAAAAAGAGAGCGGCGYGLGSQGTSGRGLGQ 430
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 542 VAGAGNGAGTGYGVNANGNGSGATGALAAVGA-----GAAGDATSGTGFG--G 594
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 431 AGAAAAAAGAGAGCGGCGYGLGSQGTSGRGLGSGGAGAAAAAAGAGAG--QGGY 489
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 595 AGGSARGLIFALGAGAA--GGDASTGVGPGCGTGTASSPFCIATAIGAGAGCGAG 652
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 490 -----GLGSQGTSGPGYGPQ--QTSRGLGSGGAGAAAAAAGAG 528
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 653 TSGATGAGGDEVFEGIAVLGLFGAGAGGAATGDCATGGAGGFGAGAGIANFLGFS 712
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 529 AAGGAGCGGCGYGLGSQGTSGRGLGSGGAGAAAAAAGAGAGCGGCGYGLGSQGTSGRG 588
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 713 VILHG--GAGGAGGTAT--GTGNGAGAGGGGGLSPVILGIGAGAGDGGGALVLG--G 768
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 589 GLGCGAGAGAA---AAAAAAGAGCGGCGYGLGSQGTSGRG--GLGCGAGAAAAA 642
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 769 GDCGGGGEAVVAGIAGVAGAGGAAPTNGAGGNGCDALIVGVGNGNAGTGTGFAN 828
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 643 AAGGAGCG-----GYGLGSQGTSG--CPGCGY-----PGQ 670
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 829 TCGNGDPTIIVNMLAPSTLIGGNGGVNGAGGTGKAGVFGAPGQ 878
```

Search completed: July 3, 2001, 14:59:40
Job time: 365 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:39 ; Search time 37.35 Seconds
(Without alignments)
633.749 Million cell updates/sec

Title: US-09-490-291-6
Sequence: 1 MASMTCGQMGKSGMASGRC.....TSGIRIRAPSTSFHHHHH 691
Perfect score: 3543

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	60.3	747	1 SPDL_NEPCL	P19837 nephila cla
2	1281.5	36.2	1901	1 YZ08_MYCTU	O53553 mycobacteri
3	1174.5	33.1	5283	1 FBOH_BOMO	P05790 bombyx mori
4	1138.5	32.1	627	1 SPD2_NEPCL	P46804 nephila cla
5	1110.5	31.3	914	1 WA22_MYCTU	O06794 mycobacteri
6	1049.5	29.6	778	1 YQ34_MYCTU	P71933 mycobacteri
7	1032	29.1	957	1 Y278_MYCTU	P56877 mycobacteri
8	988.5	27.9	801	1 Y747_MYCTU	O53810 mycobacteri
9	976	27.5	860	1 ELS_MOUSE	P54320 mus musculu
10	939	26.5	864	1 ELS_RAT	O99372 rattus norv
11	869	24.5	603	1 YD25_MYCTU	O10637 mycobacteri
12	853.5	24.1	465	1 GRP2_PHAUV	P10496 phaseolus v
13	777.5	21.9	747	1 ELS_BOVIN	P04985 bos taurus
14	763	21.5	641	1 EBN1_EBV	P03211 Epstein-Bar
15	709	20.0	750	1 ELS_CHICK	P07916 gallus gall
16	691	19.5	730	1 ELS_HUMAN	P15502 homo sapien
17	686	19.4	498	1 Y118_MYCTU	O50615 mycobacteri
18	681	19.2	1733	1 VNUN_PPKVA	P33485 pseudocari
19	678.5	19.2	543	1 YP91_MYCTU	O50603 mycobacteri
20	673	19.0	384	1 GRP1_PETHY	P09789 petunia hyb
21	669	18.9	481	1 LOR1_MOUSE	P18185 mus musculu
22	656.5	18.5	1027	1 CAFE_RLEPA	P30754 riftia pach
23	644	18.2	338	1 GRP_ARATH	P27483 arabidopsis
24	630	17.8	1373	1 CA21_MOUSE	O01149 mus musculu
25	626	17.7	515	1 Y140_MYCTU	O50594 mycobacteri
26	622	17.6	1466	1 CA13_HUMAN	P02461 homo sapien
27	616	17.4	434	1 YK98_MYCTU	O10707 mycobacteri
28	608	17.2	1418	1 CA12_HUMAN	P02421 mus musculu
29	607	17.1	1464	1 CA13_MOUSE	P08121 mus musculu
30	606.5	17.1	1453	1 CA11_CHICK	P02457 gallus gall
31	603.5	17.0	1049	1 CA13_BOVIN	P04258 bos taurus
32	602.5	17.0	463	1 YA68_MYCTU	O53416 mycobacteri
33	602.5	17.0	1460	1 CA11_CANFA	O93937 canis fami

34	600	16.9	1366	1 CA21_HUMAN	P08123 homo sapien
35	599.5	16.9	1459	1 CA12_MOUSE	P28481 mus musculu
36	599.5	16.9	1464	1 CA11_HUMAN	P02452 homo sapien
37	599	16.9	1366	1 CA21_CANFA	O46392 canis fami
38	595	16.8	1364	1 CA21_BOVIN	P02465 bos taurus
39	592.5	16.7	1453	1 CA11_MOUSE	P11087 mus musculu
40	573.5	16.2	1362	1 CA21_CHICK	P02467 gallus gall
41	563	15.8	1362	1 CA24_CAUEL	P17140 caenorhabdi
42	561.5	15.8	1567	1 ICEN_XANCT	P18127 xanthomonas
43	559	15.8	1758	1 CA14_CAUEL	P17139 caenorhabdi
44	555.5	15.7	779	1 CA11_BOVIN	P02453 bos taurus
45	553.5	15.6	1496	1 CA25_HUMAN	P05997 homo sapien

ALIGNMENTS

RESULT	1.	SPDL_NEPCL	STANDARD	PRT	747 AA.
AC	P19837				
DT	01-FEB-1991 (Rel. 17, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).				
OS	Nephila clavipes (Ord spider).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;				
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.				
OX	NCBI_Taxid=6915;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RX	MEDLINE=90384959; PubMed=2402494;				
RA	Xu M., Lewis R.V.;				
RT	"Structure of a protein superfiber: spider dragline silk.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).				
RN	[2]				
RP	SEQUENCE OF 653-747 FROM N.A.				
RX	MEDLINE=94165058; PubMed=8120021;				
RA	Beckwith R., Arcidiacono S.;				
RT	"Sequence conservation in the C-terminal region of spider silk				
RT	proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and				
RL	J. Biol. Chem. 269:6661-6663(1994)."				
CC	- FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE				
CC	CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF				
CC	PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED				
CC	WITH ELASTIC AMORPHOUS SEGMENTS.				
CC	- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.				
CC	- SUBCELLULAR LOCATION: EXTRACELLULAR.				
CC	-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M37137; AAA29380.1; -				
DR	EMBL: U03848; AAB60212.1; -				
DR	PIR: A36068; A36068.				
KW	Silk; Repeat.				
FT	NON_TER	1			
FT	DOMAIN	1	655		25 X APPROXIMATE TANDEM REPEATS.
FT	REPEAT	1	25		1.
FT	REPEAT	26	38		2.
FT	REPEAT	39	66		3.
FT	REPEAT	67	96		4.
FT	REPEAT	97	130		5.
FT	REPEAT	131	158		6.
FT	REPEAT	159	191		7.
FT	REPEAT	192	204		8.
FT	REPEAT	205	235		9.

FT	REPEAT	236	262	10.
FT	REPEAT	263	292	11.
FT	REPEAT	293	305	12.
FT	REPEAT	306	333	13.
FT	REPEAT	334	360	14.
FT	REPEAT	361	394	15.
FT	REPEAT	395	424	16.
FT	REPEAT	425	458	17.
FT	REPEAT	459	485	18.
FT	REPEAT	486	512	19.
FT	REPEAT	513	525	20.
FT	REPEAT	526	535	21.
FT	REPEAT	556	582	22.
FT	REPEAT	583	612	23.
FT	REPEAT	613	642	24.
FT	REPEAT	643	655	25.
FT	CONFLICT	662	662	V -> L (IN REF. 1).
FT	CONFLICT	672	672	S -> T (IN REF. 1).
FT	CONFLICT	695	747	NGSLSCDVALIOALLEVVASLQILGSSSICGVNYSAGQAF FOIVGOSVYOAL -> ILVFLDVMSSFKLFSRLFILLRS (IN REF. 1).
SO	SEQUENCE	747 AA:	60528 MW:	850D44BD0D49E012 CR664;

Query Match	60.3%	Score 2138	DB 1	Length 747
Best Local Similarity	66.3%	Pred. No. 3.2e-88		
Matches 498	Conservative 20	Mismatches 49	Indels 184	Gaps 35

[illegible]

Db 588 QGAG-----AAAAAGAGAGGCGTGGTGGGCGT-GRGGTGGGCGAG-----AAAAGCAAGCGGT 636

QY 653 GGLGCGCTSGPBGYGPQQTSGIRIRAPST 683

Db 637 GGVGSGASAA-----SAAASRLSSPQAS 659

RESULT	2	
Y208_MYCTU		
ID	Y208_MYCTU	STANDARD;
		PRT; 1901 AA.

DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sec

30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
P05008 OR UNP05008 1E

GN
OS
OC

RV3508 UR M1V023.12.
Mycobacterium tuberculosis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacterium
OX NCBI_TaxID=1773;
RN [1]

SEQUENCE FROM N.A.
STRAIN=H37RV;
RC

RX MEDLINE=982829398//; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horrocks T., Tatarski K., Trovati A., McLean T., Moyle S., Murphy L.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Normey P., Ouzens A., Paton D., Pearson C., Peckham G., Perkins J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.

RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
RT complete genome sequence.":
RT

RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGSR
CC CREDENTIY

CC SUBFAMILY.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isid-sib.cn.
 CC -----
 CC EMBL; AL022022; CAA17745.1; -.
 CC -----
 DR

```
DR      HSSP; P19972; 1KVD.
DR      Tuberculist; RV3508; -.
DD      InterPro: ITP0000084; -.

```

DR Pfam; PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal.

FT	SIGNAL	1	30	POTENTIAL.
FT	CHAIN	31	1901	HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT				RV3508.

SQ SEQUENCE 1901 AA; 147627 MM; C7B1923D5D0146CD CRC64;

Query Match	36.2%	Score 1281.5;	DB 1;	Length 1901;
Best Local Similarity	39.5%;	Pred. No. 1.9e-50;		

Matches 354; Conservative 28; Mismatches 273; Indels 241; Gaps 39; Overlaps 7

280	GGGCGACGACASDGGALGCTGCTGCTGGACGAGACGALLTGAGGGGGLGAGAGGCGCTGCGAG	339
-----	---	-----

QY 53 -----CTSGRGSLGG-----QGAGAAA-----AAAAAAGG-AGGGYGLGSQ 90

Db 340 GGGVGGVGGTGGKGGVGGVAGLGGAGAGAGQLFSAGCAAGAVGVGGTGGGGAGGAGAA 399

[illegible]


```

DN 01-NOV-1997 (Rel. 35, last sequence update)
DT 30-MAY-2000 (Rel. 39, last annotation update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.
GN RV2634C OR MYC1441.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parhill J., Garnier T., Churcher D.,
RA Gordon S.V., Eiglmeter K., Gas S., Barry C.E., III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajadaram M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: Z80222; CAB02341.1; -.
DR HSSP: P04002; IATE.
DR TubercuList: RV2634C; -.
DR InterPro: IPR000084; -.
DR InterPro: IPR001899; -.
DR Pfam: PF00934; Pf: 1.
DR KW Hypothetical protein.
SQ
SEQUENCE 778 AA: 63131 MW: DAB20FE58E4999E7 CRC64:

Query Match 29.6%; Score 1049.5; DB 1; Length 778;
Best Local Similarity 41.3%; Pred. No. 1.8e-40;
Matches 291; Conservative 21; Mismatches 294; Indels 99; Gaps 34.

OY 7 GQOMGRGSAASRGRLGCGGAGCAAAAAAAGAGAGGCGGCGTSG-----GTSGR 57
DB 129 GARGGGGGLLTNGGNGSGSAPQPCGAGDGLTNGGTGGCGDGLVSGAAGVGGR 188
OY 58 GG-TLGGGAGAAAAAAGAGAGGCGGCGTSGRGLGCGAGAAAAAAG 115
DB 189 GGHLLNGGTGAGGAAGATLVG-GRGVG--GATLLISGGRG--GAGGAAAGVGTGG 243
OY 116 GAGGCGGGLGSGQTSGRGLGCGGAGAAAAAAGAGAGGCGGCGTSGRPGCIG 175
DB 244 VGSSSGVGV--FGNGCGFGAGGLGAAGGVGAASYFTGTGGGGVGDGAPG--CGGAG 299
OY 176 PGOQTSGRGLGCGGAGAAAAAAGAGAGGCGGCGTSGRGLGCGGAGAAAAA 235
DB 300 P--LLTNGNGVGLG-GAGAA-----GGNGGAGGM-----LLGGGAGGCGGPAVAGV 344
OY 236 AAAAAGAGAG-----GYGGLSGQTSGRHGLGCGGAGAAAAAAG-----GAG 283
DB 345 LGGMPAGGAGGGAANMFSGSAGAGGGGTGLAAGTGNVPSIAPNTGANGCTDMSGNCOT 404
OY 284 GGYGGLSGQTSGR-GGLGCGAGAAAAAAGAGAGGCGGCGTSGRPGYGPQ 342
DB 405 GGGGCGPAGCGVAGAGVGSGG-----GLGESLDGNDGTGGKGGAG--GTAATDG-GAG- 455
OY 343 QTSGRGGLGCGAGAAAAAAGAGAGGCGGCGTSGRGLGCGGAGAAAAA 402

```

[illegible]

RESULT	7
Y278_MYCCTU	STANDARD; PRT; 957 AA.
ID	Y278_MYCCTU
AC	P56877;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	HYPOTHEITICAL PE-PGRS FAMILY PROTEIN RV0278C. PRECURSOR.
OS	RV0278C OR MTW035.06C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE=96295987; Pubmed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Hornsbury T., Jurgels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutler S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrell B.G.;
RT	Deciphering the biology of Mycobacterium tuberculosis from the
RT	complete genome sequence.;
RL	Nature 393:537-544(1998).
CC	-1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC	SUBFAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; AL021930; CAA17353.1; -.
DR	TuberculoList; RV0278C; -.
DR	InterPro; IPR000084; -.
DR	pfam; PF00934; PE; 1.
KM	Hypothetical protein; Repeat; Signal.
FT	SIGNAL 1 30
FT	CHAIN 31 957
FT	POTENTIAL
FT	HYPOTHEITICAL PE-PGRS FAMILY PROTEIN
FT	RV0278C.
SO	SEQUENCE 957 AA; 81905 MW; 71EBAD417FBA47C CRC64;

Db 578 GCAGGGSTTAGGACGACGAGNG-----GTFAGAGTGGAGSHSTAAAGVGGAGGA--G 6253

QY 487 GYGGLGSGQGTSGPEGYGPBQQTSGRGGLGCGAGAAAAAAGAGGCGYGGLSQGT 546

Db 626 GDADLLSLGASGGAGGSGSGSLTAAGYVVGIG----GAGGLLTFGSGGAGSGGFGNSGN 680

QY 547 SGRGGLGGQ-----GAGAAAAAAGGAGGAGGCGYGGLSQGTSGRGGLGCGAGAAAA 601

Db 681 GCAGGAGGDCAGLLVSSGAGGAGCAGASAT--GAATGGDGGAG--GKSGAFGLGGDGGAGGATG 737

QY 602 AAAAAAGAGGCGYGGLSQGTSGRGGLGCGAGAA--AAAAAAGAGGAGGCGYGG--LGSQ 658

Db 738 LSGAFHITGGKGGVG--GSAVLIGNGGNGGNGNGNAGKSGGAPGPGAGGAGGALLGEN 795

QY 659 GTSG 662

Db 796 GLNG 799

RESULT	9			
ELS_MOUSE	ELS_MOUSE	STANDARD:	PRT:	860 AA.
AC	P54320;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	ELASTIN PRECURSOR (TROPOLASTIN).			
GN	ELN.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C; TISSUE=Lung;			
RX	MEDLINE=95130069; PubMed=7829060;			
RA	Wagner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;			
RT	"Use of an Inlition Polymorphism to Localize the Tropoelastin Gene to			
RT	mouse chromosome 5 in a region of linkage conservation with human			
RT	chromosome 7.";			
RL	Genomics 23:125-131(1994).			
CC	-1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND			
CC	NIDHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.			
CC	-1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER			
CC	INTO AN EXTENSIBLE 3D NETWORK.			
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.			
CC	-1- PFM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-ch).			
CC	or send an email to license@isb-sib.ch).			
CC				
DR	EMBL; U08210; AAA80155.1; -			
DR	HSSP; P04002; IWPA.			
DR	MED; MG1:95317; Eln.			
FT	Structural protein; Repeat; Signal; Connective tissue.			
FT	SIGNAL			
FT	CHAIN			
FT	1 27			
FT	POTENTIAL.			
FT	ELASTIN.			
FT	28 860			
FT	SEQUENCE 860 AA; 71955 MW; 0C0BE5AE1EDD7F1 CRC64;			

Query Match	27.5%;	Score 976;	DB 1;	Length 860;
Best Local Similarity	39.38;	Pred. No. 3.1e-37;		
Matches 323;	Conservative 36;	Mismatches 249;	Indels 214;	Gaps 40;

```

0y 13 GSMASGRGIGGG-----AGAAAAAAAAAGAGCGCGGG 48
    | : : | | | | | | | | | | | | | | | | | |
Db 78 GTPGAGPGGIGGAGPAGLGAFPAGTTPGAGALYPGGAAGAAAYKAAAKAGAGLGGVG 137

```

Oy	49	LGSGCTSGRGL- -GGCGAGAAAAAAGAGAGGCGGGLGSGTSGRGLGCGAGAAA	107
Db	138	V--PGGVGGVPGGVG-- - - - - -GVPGGVGGVPG-- - - -GVGGTGGTGGVSTGA	184
Oy	108	AAAAAAG-GAG--GGGCGTSGSGTSGRGLGCGA- - - - -GAAAAAANA	151
Db	185	VVPQVAGTIGAGCRPKGYPTGVLPTGTARFPFGVGLPFGVPTGTGKAKAPG	244
Oy	152	AGGAGCG--GYGGLGSG- - - - -GTSRPGGY- - - - -PQQTSGRGLGCG- - - - -	189
Db	245	GCGAFSGCIPGCGPFGGQDPGVPLGPIKAPKPLPGCYGLPTYNGLPLPGVAGAGAKAGPYT	304
Oy	190	--GAGAAAAAANAAGAGGCGGY- - - - -GLSGGTSGR- - - - -GGTGGGAGAAAAA	237
Db	305	GTVGSGQAAAAAARAKYGAGAGVLPBGVGGGIPGAGAIPTGIGTAGAGTPAAAAAAK	354
Oy	238	AAAGGA- - - - - - - - - - -GGGCGGCGTSGRGLGCGAG	268
Db	365	AAAKRAKGAAGGLVPGGPRVRLPGAGTIPGVGGTIPGVGLIPYGVGPIIGRPGLVGGAV	424
Oy	269	AAAAAANAAGGA- - - - -GGGGY- - - - -GLGSGTSGRGLGCGGAGAAAAAAGAGCG	321
Db	425	SPAAAKRAAARAKYGARGGVGIVTYGVGAGGFPFGYGVGAGAGLGGASPAAAAAAARAK	484
Oy	322	GGYGGTSGGTSRGG- - - - -GYRPG- - - - -QOTSRRGLGCGGAGAAAAAANA- - - - -AG-GAG	370
Db	485	YGAGGAGLGLVPGAVPALPGVPAVPVPGGVPVPGAGTPTAAAAAANAARAKAGLPG	544
Oy	371	QGGY-GGLGSGTSGRGLGCGGAGAAAAAANAAGAGAGGCGGCGTSGTSGRGLGCG	429
Db	545	VGVPGCGVGVGITGCGVGVG- - - - -VPGGVPRGCVTGTG- - - - -ACRGLGGA	588
Oy	430	GAGAAAAAANA- - - - -AAA-GAGGCGYG-GLGSGTSGRGLGCGGAGAAAAAANA	478
Db	589	GSPAAKRAAARAKQYRAAAGLGGAGVPGCGAGAGVPGCGAGGVPGFAGAGVPGFCA	648
Oy	479	AAG- - - - -GAG- - - - -QGGYGGTSG- - - - -GTSRPGGYRPGQOTSRRGLGCGG- - - - -	518
Db	649	GAGVPGFAGAVPGLSLAARAKYGAAGLGGPGLGPGGLG- - - - -GPGTGGGAVP	702
Oy	519	--AGAA-AAAAAAGAGAGCGGGGGLGSGTSGRGLGCGAGAAAAAANAAGAGAGCG	574
Db	703	GRVAGAAAPPAANAARAKRAAAYDTGLGAGAGLGGAGLGAAGGL- - - - -AGTGLAG	753
Oy	575	GY--GGLGSGTSGRGLG-GCGAGAAAAAANAAGAGAGGCGGYGGLGSGTSGRGLGCG	631
Db	754	GLGAGGLGAGGL-GAGGLGAGAGGVSPAAAKAAKYGAAGLGV--LCARFPGGVAARP	810
Oy	632	GAGAAAAAANAAGAGAGCG- - - - -YGG-LGSGTSGRGGY	667
Db	811	GFGTSPITPGGGAGLGVGCGPPKPYGGALGALGTYOGGCGFG	852

	RESULT	10
ELI5_RAT		
ID	ELI5_RAT	STANDARD; PRT; 864 AA.
AC	Q99372.	
DC	01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	01-OCT-2000 (Rel. 40, Last annotation update)	
DE	ELASTIN PRECURSOR (TROPOLASTIN) (FRAGMENT).	
GN	ELN.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=91104868; PubMed=1702999;	
RA	Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;	
RT	"Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";	
RL	Biochemistry 29:9677-9683(1990).	
RN	[2]	

RP MEDLINE-781-864 FROM N.A.
RX MEDLINE-68330868; PubMed-2971041;
RA Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA."
RJ J. Biol. Chem. 263:13504-13507(1988).
RN [3]
RP SEQUENCE OF 264-533 AND 558-864 FROM N.A.
RX MEDLINE-92241859; PubMed-1572637;
RA Pierce R.A., Alatalo A., Deak S.B., Boyd C.D.;
RT "Elements of the rat tropoelastin gene associated with alternative
RT splicing."
RL Genomics 12:651-658(1992).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NOCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -1- ALTERNATIVE PRODUCTS: THREE DIFFERENT MRNAS HAVE BEEN FOUND THAT
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PFM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60647; AAA42269.1; -;
DR EMBL; J04035; AAA42268.1; -;
DR EMBL; M86372; AAA42271.1; -;
DR EMBL; M86355; AAA42271.1; JOINED.
DR EMBL; M86363; AAA42271.1; JOINED.
DR EMBL; M86364; AAA42271.1; JOINED.
DR EMBL; M86366; AAA42271.1; JOINED.
DR EMBL; M86371; AAA42271.1; JOINED.
DR EMBL; M86376; AAA42272.1; -;
DR EMBL; M86373; AAA42272.1; JOINED.
DR EMBL; M86375; AAA42272.1; JOINED.
DR HSSP; P04002; IWFA.
KW Structural protein; Connective tissue; Repeat; Signal;
KW Alternative splicing.
FT NON_TER 1
FT SIGNAL 21
FT CHAIN 22
FT VARIAT 263 307
FT VARIAT 308 308
FT VARIAT 809 823
FT SEQUENCE 864 AA: 72786 MW: 456894BB09E79FD4 CRC64;
BY SIMILARITY.

Query Match	26.5%	Score 939	DB 1	Length 864
Best Local Similarly	38.3%	Pred. No. 1.3e-35		
Matches 316	Conservative 32	Mismatches 279	Indels 198	Gaps 38

Qy	20	GGGCGGAGGAAAAAAAAG-----GAGGGGAGGGLGSGG-----TSGRGG--LGGGGAGAA	68
Db	53	GGGCGGALGPEGGRPPKRGAGILLGAFGAPGGGLGAGCGAGLSTASRRGGVLYVPGGGAGAA	112
Qy	69	AAAAAAGAGAGGGGAGGGL-----GSGG-----TSGRGGGCGGAGAAAAAAA	111
Db	113	AAATAAAKAGAGGCGGIGGIPGGVGGVGPAGVGGVAGVGGIGGIGGIGVSTGAVVPG	172
Qy	112	AAAG-GAG--GGGCGGGLGSGGTSRGGGLGCGA-----GAAAAAAAAGCA	155
Db	173	LGAGVGGAGGRGKVPGVGLPGVYPGVGLPGTGAFFPVGVLPGVPTGTGAKVPGGGGG	232
Qy	156	GGGGCGGLGSGGTSG-----PGGYS-----PGQOTSGRGLGCG-----G	190
Db	233	AFSGIIPGVGPFPGCGGQPGVPLGPIYTKAKRLPGGYGLPPTTNGKRLPRGVAGAGGKAGYPRGTG	292
Qy	191	AGAAAAAAGGAGGAGGCG-----GYGGLGSGGTSGR-----GGGCGGAGAAAAAAA	239

Dd	293	VGSGAAVAAAAAAYAGAGGGGCVLPVYGGGGIPGGAGAPETGGITGAGPPAAAAA	352
Oy	240	AGGAGGGGCGGLGSGGTISGR--GGLGGCGAGAAAAAAAAMG--AGCGGCGGLSGGTISGR	297
Dd	353	AKAAKYYGAAGGCLVPGPGPVRVPVAGIIPDVGIPGVGIPGVGIPGVGIPGVGPGTGGP	412
Oy	298	GGLGGCGAGAAAAAAAAGAGGGGGGGLG--SGGTISPGGYPGGOOTSGRGGLGG--Q	353
Dd	413	GIVGPGPVPSPAAAAKAAKAAKAAKGAARGVGIPITYGV--GAGGF--PGYGVGAGAGLGSASQ	470
Oy	354	GAGAAAAAAAAGAGAGGGGCGGCGI-----GSGTISGRG	386
Dd	471	AAAAAAAKAAKYGAGGAGTGLGVLPGAVPALPGAVPALPGAVPALPGAVPGVPTG	530
Oy	387	GLGGCGAGAAAAAAA-----AGAGGGCG--GGLGSGGTISG-----	422
Dd	531	GVPGAGTPAAAAAAAAGAAAAAKAGQYGLGPGVGVPGGVGVGGILPBGVPGGVGIGTGP	590
Oy	423	-----RGGLGGCGAGAAAAA-----AAAG--GAGGGGCGY-----GLGSGGTIS	459
Dd	521	GTGVLPGDGLGGAGTPAAAKSAKAAKAAKQYHAAAGLAGVGVGLGVAGVPBGAGAGGFC	650
Oy	460	GRGGLGSGGAGA-----AAAAAAAAGAGAGGGGCGGGLGSG--GTSPPGCGYGPBQOOTSGRG	513
Dd	651	AGAGYPCGAGAVPBGSLAASRAAKYGAAGGGLGCGPGGLGCGPGGLGPGGFC-----GPGG	704
Oy	514	LGGGAGAGAA-----AAAAAAAAGAGAGGGGCGGGLGSGGTISGRGGLGGGAGAAAAAAAAGA	571
Dd	705	LGGVPGVGAAGAPAAAAAAKAAKAAKRAOTGILGGAGGGLGAGLGAAGL-----AGGL	755
Oy	572	GGGCGY--GGGLSGOGT--SGRGGLGAGGAGAAAAAAAAGAGAGGGGCGGGLGSGGTISGRGGL	628
Dd	756	GAGGLGAGGGLGAGGVTPEAVGLG--GVSPAAAAKAAKYGAGLGGV--LGRAPPPGGVA	811
Oy	629	GGCGAGAAAAAAAAGAGAGGG-----YGG--LGSOGTISPPGCGTG	667
Dd	812	ARPPGGLSPITPGGAGGGLGVGAGPPKPYGTGALCALGATGGGGCGTG	856

RESULT 11

ID	YD25_MYCTU	STANDARD;	PRT;	603 AA
----	------------	-----------	------	--------

AC Q10637;

DT 01-OCT-1996 (Rel. 34, Created)

DT	30-MAY-2000 (Rel. 39, Last annotation update)
DI	01-JUL-1996 (Rel. 34, Last sequence update)

DE HYPOTHETICAL, PE-PGBS FAMILY PROTEIN BV1325C E

GN RV1325C OR MTCY130.10C.

05 Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinob

NCBI TaxID=1773: Actinomycetales; corynebacterineae; Mycobacte

```

D 0
RN [1]
mod-10000-1/10

```

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

Colo's E. Broch. P. Parkhill T. Carrfor E.
MEDLINE=98295987; PubMed=9634230;
RX

BA Gordon S.Y., Eidlmeier K., Gas S., Barry C.E.,
KA Cole S.L., Broccoli N., Falkner D., Gannier J.

RA Badcock K., Basham D., Brown D., Chillingworth

RA Davies R., Devlin K., Feltwell T., Gentles S.

RA Hornsby T., Jagels K., Krogh A., McLean J., M

RA Oliver S., Osborne J., Quail M.A., Rajanandran
RA Butler S., Seeger K., Skelton S., Squares S.

RA Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of *Mycobacterium tuberculosis*"

RT complete genome sequence. #;

RL Nature 393:53/-344 (1998).

1. SIMILARITY: BELONGS TO THE PROCEEDING FAMILY

CC SUBFAMILY.

CC

CC

.....

CC This SWISS-PROT entry is copyright. It is protected by the copyright law of Switzerland.

CC between the Swiss Institute of Bioinformatics

[illegible]


```

Db 297 APCAIPGIGIAGVAGAPDPAAAAAAAAAAKFGAAGLPGVGVPPVGVPGVPGVVP- 355
Oy 310 AAAAAAAAAAGAGGGG--GLGSGGTSGPGGVPGGQOTSGRGGLGQGAAGAAAAA- 366
Db 366 -----GAGVPGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV- 407
Oy 367 -GGAGGGGGLGSGGTSGRGGLGSGG-AGAAAAAAGAGAGGGGGLGGS----- 417
Db 408 FGARGAVGIGIPTFGL-GPGGFPGIDGAAAAAATAAKKAGAGGALGVVPGAPG 466
Oy 418 --QGTSRGGLGQGAAGAAAAAG--GAGG-----GCGGLGSGGT-- 459
Db 467 AIPGLPGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV- 526
Oy 460 ---GGGGLGAGAGAAAAA-----GGAG--GGGGLGSGGTSGPG--GYPGQOT 508
Db 527 IGLGPGGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV- 585
Oy 509 SGRG-GLGGGGA--GAAGAAAAAAGAGAGGGGGLGSGGTSGRGGLGQGAAGAAAAA 565
Db 586 LGVAGAGVPGVAVPGTAAAKAKKFGPGGVGALGGVGDGCGAGITG--GVAAGVPAAGAA 644
Oy 566 AAGAGAGGGGGLGSGGTSGRGGLGCG--GAGAAAAAAGAGAGGGGGLGSGG 621
Db 645 KAAAKAAGGGLGAGGGLGAGVAVGAGVAGVAGVAVAGVAVAGVAVAGVAVAGV- 704
Oy 622 TSGRGGLGSGGAGAAAAAAGAGAGGGGGLGSGGTSGPGG 665
Db 705 PLG-GGAGGLG-----VGGRPKRPFPGALGALGF--PGG 735

```

RESULT 14

```

ID EBN1_EBV STANDARD: PRT: 641 AA.
AC P03211:
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE EBN1-1 NUCLEAR PROTEIN.
GN BKRF1.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptoviruses.
OC NCBI_TaxID=10377;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuftnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RT Nature 310:207-211(1984).
RN [2]
RP SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE=86259739; PubMed=3460083;
RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RT "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RT proteins: a probable transcriptional initiation site.";
RT Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=90266473; PubMed=2161150;
RA Petit L., Sample C., Kieff E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT latent infection nuclear proteins.";
RT Virology 176:563-574(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX MEDLINE=96006523; PubMed=7553871;
RA Bocharov A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,
RA Edwards A.M., Trappier L.;
RT "Crystal structure of the DNA-binding domain of the Epstein-Barr

```

```

RT virus origin-binding protein EBNA 1.";
RL Cell 83:38-46(1995).
CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSCRIPTIVATING FACTOR
CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOLUS, SOMEWHAT
CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC WITH THE NUCLEAR MATRIX.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; CA24816.1; -.
DR EMBL: M13941; AA45889.1; -.
DR PIR: A03773; Q0BE31.
DR PIR: S33021; S33021.
DR PDB: 1VHI; 23-DEC-96.
DR TRASNSEC: T00211; -.
KW Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW 3D-structure.
FT DOMAIN 87 352 GLY/ALA-RICH.
SQ SEQUENCE 641 AA; 56427 MW; 4D161653E16FC341 CRC64;

```

Query Match 21.5%; Score 763; DB 1; Length 641;

Best Local Similarity 35.7%; Pred. No. 5,8e-28;

Matches 232; Conservative 22; Mismatches 209; Indels 186; Gaps 24;

```

Oy 116 GAGGGGGLGSGG--TSGRGGLGAGGAAAAAAGAGAGGGGGLGSGGTSGPGY 174
Db 5 GPGTGNGLGKCDTSGEGSGSPQRRGDNHGRGRGRGRGCG--GRGAPGSGSS 62
Oy 175 GPGGQTSR-----GGLGGGAGAAAAAAGAGAGGGGGLGSGGTSGRG 222
Db 63 GPHHRGVRRPQKPSCTGCKTHGGTGAG--AGAGAGAGAGAGAGGAGAG--GAGGAG 119
Oy 223 GLGGGAGAAAAAAGAGAGGGGGLGSGGTSGRGGLGSGGAAAAAAGAG 282
Db 120 GAGGAGAG-----GGAGAG-----GAGAGAGAGAG--GAGAGGGAGAG 158
Oy 283 GGGYGLGSGGTSGRGGLGAGAAAAAAGAGAGGGGGLGSGGTSGPGYGPQ 342
Db 159 AGG-----GAGGAGAGAG-----GGAGAG-----GAGGAGAG-- 187
Oy 343 QTSRGGLGAGGAAAAAAGAGAGGGGGLGSGGTSGRGGLGSGGAAAAA 402
Db 188 --GGAGAGAGAG-----GGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 235
Oy 403 AAGAGAGGGGGLGSGGTSGRGGLGAGAAAAAAGAGAGGGGGLGSGGTSGRG 462
Db 236 GAGGAGAG-----AGAGAGAGAGAGAGAGAGAGAGAGAGAG-----GAGGAGAG 282
Oy 463 GLGGGAGAAAAAAGAGAGGGGGLGSGGTSGPGYGPQTSRGGLGAGAG 522
Db 283 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
Oy 523 AAAAAAAGAGAGGGGGLGSGGTSGRGGLGAGGGA--GAAGAAAAAAGAGAGGGGGLG 581
Db 334 -----SGRGAGGGGAGAGGGGGRGRGRGRGRGRGRGRGRGRGRGRGRGR 386
Oy 582 QGTS-----GR-----GGLGAGGAGAAAAAAGAGAGGG 613
Db 387 QSSSSGPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPP 446
Oy 614 YGGLGSGGTSGRGGLG-----GAGAGAAAAAAGAG-----AG 648
Db 447 STGPRGQDGGRRKKGWPKHKGQGSNPKENIEGLRALLARSHVERTTDEGTWAG 506

```


OY	122	YGSLSGSOGTSGRGLGCGAGAAAAA	AAAAAAAAAAAGAGAGCGGTGGTSCGTSGPGCTGPCQQTSS	181
OY	122	: :	: :	181
Dd	101	YGGJLSQG-AGRGSQGGAC-	-----AAAAAACGAGCGGTGGLSQG-----A	142
OY	182	GRGJLGGCAGAAAAAAAAG-	-----GAGDGYGGTCSQGTSGRGLGCGCAGA	231
OY	182	: :	: :	231
Dd	143	GRGJLGGCAGAAAAAAGAGCGGTGGTSGCGGTGGTSSQG-AGRGLGCGAG-	200	
OY	232	AAAAAAAAAGAGCGGTGGTCSQGTSGRGLGCGCAGAAAAA	AAAAAGAGCGGTGGTCS	291
OY	232	: :	: :	291
Dd	201	-AAAAAAGCAGC-	-----GGTGGGAGCAGCAAAAAAGCAGCGGTGGTCS	244
OY	292	OGTSGRGLGCGCAGAAAAA	AAAAAGCAGCGGTGGTCSQGTSGPGCTGPCP-GOOTSRRGL	350
Dd	245	OG---AGRGQGGAGAAAAAV--	--GAGCGGTGGC---AGCGGTGGTCSQCGRGRGL	292
OY	351	GCGCAGAAAAAAGAGCAGCGGTGGTCSQGTSGRGLGCGCAGAAAAA	AAAAAGCAGCG	410
Dd	293	GCGCAG---AAAAAAGCAG-	-----GGTGGGAGCAGCAAAAAAGCAGCG	335
OY	411	GYGJLGSQGTSGRGLGCGCAGAAAAA	AAAAAGCAGCGGTGGTCSQGTSGRGLGCGCAG	470
Dd	336	GYGJLGNQG-AGRGQG-	-----AAAAAAGCAGCGGTGGTCSQG-AGRCGTGGCAG	384
OY	471	AAAAAAAAAAGCAGCGGTGGTCSQGTSGPGCTGPCP-GOOTSRRGLGCGCAGAAAAA	AAAAA	529
Dd	385	----AAAAAAGCAGCGGTGGTCSQGTSGRGLGCGCAG---	AAAAA	435
OY	530	AGCAGCGGTGGTSGRGLGCGCAGAAAAA	AAAAAGCAGCGGTGGTCSQGTSGRCG	588
Dd	436	AGCAGC-----GTLGCGCAGCGCAGAAAAAGCAGVRGCRTGGTCSQ---	AG	477
OY	590	LGGCAGAAAAAAGAGCAGCGGTGGTCSQGTSGRGLGCGCAGAAAAA	AAAAAGCAG	649
Dd	478	RCCGAG-----AAAAAAGCAGCGGTGGTCSQGTSGRGLGCGCAG---	AAAAAGCAG	526
OY	650	GCGJLGSQGTSGPGCTGPCPOOTSIRRASTSS	683	
Dd	527	GCGJLGSQGTSGPGCTGPCPOOTSIRRASTSS	552	
RESULT	2			
ID	046171	PRELIMINARY;	PRT; 544 AA.	
AC	046171;			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	SPIDROIN_1 (FRAGMENT).			
OS	Eukarya clavipes (Ord spider).			
OC	Nephilys; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Nephilys; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OX	NCBI_TaxId=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90384959; PubMed=2402494;			
RT	Xu M., Lewis R.V.;			
RL	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 449-544 FROM N.A.			
RX	MEDLINE=98148687; PubMed=9487707;			
RT	Arcidiano S., Melio C., Kaplan D., Cheley S., Bayley H.;			
RT	"Purification and characterization of recombinant spider silk			
RT	expressed in Escherichia coli.";			
RL	Appl. Microbiol. Biotechnol. 49:31-38(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	Beckwith R., Arcidiano S., Stote R.;			
RL	Insect Biochem. Mol. Biol. 0:0-0(1998).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RP	Beckwith R.;			

RL Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
 R# EMBL: U20329; AAC38957.1; -.
 FT NON_TER 1
 SQ SEQUENCE 544 AA; 44107 MW; CC611B3551945615 CRC64;

```

Query Match          48.8%; Score 1729.5; DB 5; Length 544;
Best Local Similarity 61.3%; Pred. No. 1,4e-105;
Matches 409; Conservative 12; Mismatches 33; Indels 213; Gaps 28;

QY      18 GRGGLGGCGAG-AAAAAAAAAAGCAGCGCYGGLGSQGTSGRGLGGCGAGAAAAAAA 76
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1  GYGGLGGCGAGCGAGAAAAAAGCAGCGCYGGLGSQG-----AGRGGCGAG----- 52

QY      77 GGAGCGCTGGTGLSQGTSGRGGLGGCGAGAAAAAAAAGCAGCGCYGGLGSQGTSGRGRL 136
       7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      53 GGAGCGCTGGTGLSQGTSGRGGLGGCGAG-----AAAGAGGYGQ-----GGL 92

QY      137 GGCGGCAAAAAAAGAGAGCGCYGGLGSQGTSPGCGPCGCGTSGRGGLGGCGAGAAAA 196
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      93 GGCGGCGAGAGAAAAAGCAGCGCYGGLGSQG-----AGRGSSGCGAG----- 135

QY      197 AAAAAAGCAGCGCYGGLGSQGTSGRGGLGGCGAGAAAAAAAAGCAGCGCYGGLGSQGT 256
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      136 AAAAAAGCAGCGCYGGLGSQG -AGRGGLGGCGAG---AAAAAAGCAGCGCYGGLGCG-- 189

QY      257 SGRGGLGGCGAGAAAAAAGCAGCGCYGGLGSQGTSGRGGLGGCGAGAAAAAAA 316
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      190 -----GAGCGGYYGGLGSQG -AGRGGLGGCGAG-----AAAAA 220

QY      317 GGAGCGCYGGLGSQGTSPGCGYGPQGTSGRGGLGGCGAGAAAAAAAAGCAGCGCYG 376
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      221 GGAGCG-----GGLGGCGAG-----AAAAAAGCAGCGCYG 250

QY      377 LGSQGTSGRGGLGGCGAGAAAAAAGCAGCGCYGGLGSQGTSGRGGLGGCGAGAAAA 436
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      251 LGSQG----AGRGCGCGAG-----AAAAAGCAGCGCYG----- 280

QY      437 AAAAAAGCAGCGCYGGLGSQGTSGRGGLGGCGAGAAAAAAAAGCAGCGCYGGLGSQGT 496
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      281 -----OGAGCGYYGGLGSQG -AGRGGLGGCGAG-----AAAAAAGCAGCG-- 320

QY      497 SGPGYGPQGTSGRGGLGGCGAGAAAAAAGCAGCGCYGGLGSQGTSGRGGLGGCG 556
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      321 -----GGLGGCGAG-----AAAAAGCAGCGCYGCGCA-----GGG 351

QY      557 AGAAAAAAGCAGCGCYGGLGSQGTSGRGGLGGCGAGAAAAAAAAGCAGCGCYG 616
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      352 AGAAAAAAGCAGCGCYGGLGSQG-----AGRGCGCGAG-----AAAAAAGCAGCGCYG 403

QY      617 LGSQGTSGRGGLGGCGAGAAAAAAGCAGCGCYGGLGSQGTSPGCGYGPQGTSGIR 676
       1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      404 LGCGG-GRGGLGGCGAGAAAV-----GAAGGYYGVSGASAA-----SMAAR 448

QY      677 TRAPSTS 683
       : : | |
Db      449 LSSPAS 455
```

RESULT 3
 ID 017434 PRELIMINARY; PRT: 988 AA.

AC 017434;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
 DE MINOR AMPULATE SILK PROTEIN MSPL (FRAGMENT).
 OS Nephila clavipes (Ord spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
 OX NCBI_TaxID=6915;
 RN [1]
 RP SEQUENCE FROM N.A.


```
Db 1348 SAAAAAAAAAASGAGSGGGYGMGGYGSDSAAAAAAAAASGAGSGGYGC 1407
QY 579 LGSQGT-----SGRGGLG-----OGAGAAAAAAAAAGAGAGQG 612
Db 1408 YGSDSAAAAAAAAAASGAGYGVGGYGMDDGSGSDSAAAAAAAAAASGAGSGS 1467
QY 613 GYGGIGS-----GCTSGRGGLGCGAGCAAAAAAAAA-----AGGAG 648
Db 1468 GYGGYGSDSAAAAAAAAAASGAGAGGYGGYGSDSAAAAAAAAAASGAGGAG 1527
QY 649 GGGY-----GGIGS-----OGTSGPGGYPGQOTSGTRIRAPSTFEHH 687
Db 1528 -GGYGMGGGCGYGSAAAAAAAAAASGAGGAGGDDGCGTSSSSAAAAAAAAARRAG 1586
QY 688 H 688
Db 1587 H 1587

RESULT 5
053559 PRELIMINARY; PRT; 1489 AA.
AC 053559:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PGRS-FAMILY PROTEIN.
GN RV3514 OR MTV023.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OK NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Kiroh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL, AL022022, CAI17751.1; -.
DR TubercuList; RV3514; -.
DR InterPro; IPR000084; -.
DR InterPro; IPR002173; -.
DR Pfam; PF00934; PE. 1.
DR ProDom; PD001223; -. 1.
DR PROSITE; PS00583; PKRB_KINASES_1; UNKNOWN_2.
SQ SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3CBAF3A CRC64;
```

```
QY 132 GRGIGGCGA-----GAAAAAAAAAGGA-----GGGCGTGLSQGTSGPGYG 175
Db 392 GGGAGGAGADADPGATGCTGFAGAGAGAGAGSSGAGGTNGSGGAGGAGAGGA 451
QY 176 PGQOTSGRGGLGCGAGAAAAAAAAAAGAGGCGYGC-LGSGGTSGRGGLGQ----- 227
Db 432 GADNFTGIGTGCGDGTGGAAGAGAGGAAGTGTGCMIGTTGNAGYGVGAGGCGGDDGAG 511
QY 228 GAGAAA-----AAAAAAGGA-----GQGGYGLGSGQ----- 255
Db 512 GAGADADPGATGCTGFAGAGAGAGAGAGSSGAGGTNGSGGAGGTGGGAGAGAGAD 571
QY 256 -TSGRGGLGCGAGAAAAAAAAAAGAGAGCGYGC-LGSGGTSGRGGLGQ-----GAGAAA 309
Db 572 NPTGIGTGCGDGTGGAAGAGAGAGGAAGTGTGCMIGTTGNAGYGVGAGGCGGDDGAGAG 631
QY 310 AAA-----AAAAGAGQ-----GGYGLGSGQ-GTSGPGYPGQOT 344
Db 632 ADADPGATGCTGFAGAGAGAGAGAGSSGAGGTNGSGGAGGTGGGAGAGADNPT 691
QY 345 SGRGGLGCGAGAAAAAAAAAAGAGGCGYGC-LGSGGTSGRGGLGCGAGAAAAAAAA 403
Db 692 -GIGGTGDDGCTGGAAGAGAGAGAGAGTGTGCMIGTTGNAGYGVGAGGSSGAGGTNGSGGA 750
QY 404 AGGAGCGYGLGSGQGT-----SGRGGLGCGAGAAAAAAAAAAGAGGCGYGC-LGSGGTS 459
Db 751 GCTDGGGAGGAGAGADNPTGTGTGDDGTGGAAGAGAGAGAGTGTGCMIGTTGNA 810
QY 460 GRGGLGCG-----GAGAAAAA-----AAAAAGGA----- 483
Db 811 GYGGAGCGGCGGAGAGAGADADPGATGCTGFAGAGAGAGAGSSGAGGTNGSGAGGT 870
QY 484 -----GQGGYGLGSGQTSPPG-YGPGQOTSGRGGLGCGAGAAA 523
Db 871 CGQVYAGAGAGISFNSNGSNGCTGTGCGY--GTGCGGAGNAGTACDPKRGGTGTGTG 927
QY 524 AAAAAAAAAAGAG-QGCTYGLGSGGTSGRGGLGCGAGAAAAAAAAAAGAGCGYGLGSGQ 582
Db 928 SCGAGCGGAGANFNGTGTGTG--GTGKGGLMTDGLSSATSGTGTGTGCGKGTGAGDD 985
QY 583 GTSGRGGLGCGAGAAAAAAAAAAGGA-----GQGGYGLGSGQTSRGGLGCGQ----- 631
Db 986 SAGGTGCTGTGAGAGNAGAGLANTGTAGNAGTGTGCGGCGGAGGCGGDDSG-SGLGCGPFRAG 1044
QY 632 -----GAGAAAAA-----AAAAAGAGCGYGLG-----SGGTSGPGYG 667
Db 1045 GAGGKAGAGSSGAGGTNGSGAGAGAGGCGGAGAGISFNSNGSNGCTGTGCGVG 1098

RESULT 6
053557 PRELIMINARY; PRT; 1079 AA.
AC 053557:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PGRS-FAMILY PROTEIN (FRAGMENT).
GN RV3512 OR MTV023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OK NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Kiroh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
```

RA Taylor K., Whitehead S., Barrell B.G.:
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 Nature 393:537-544(1998).
 RL EMBL: AL022022; CAI17749.1; -
 DR TubercList; RV3512; -
 DR InterPro: IPR002202; -
 DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 1079 AA; 81163 MW; A79718CDB74B97D CRC64;

Query Match 35.1%; Score 1242; DB 2; Length 1079;
 Best local similarity 40.7%; Pred. No. 1,1e-73;
 Matches 316; Conservative 34; Mismatches 306; Indels 120; Gaps 29;

OY 5 TGGGOMRGSMASGRGLGCGAGAAAAAAGA-GGCGYGC-----LGSQG 53
 DB 58 TGGGQGNCGN--GGNGGTGKGTGCGDGLAGSSGAGKKGCGDAGKAGTGSAPGTAG 115 -
 OY 54 TSGRGLGCGG-----AGAAAAAAGA-GGCGYGC-----LGSQG 101
 DB 116 TGGDGGKGGNGGICAAAGTTPVGTGASGCTGCGAGGTGGDGAANGCTAGAGAGAGNG 175
 OY 102 -----GAGAAAAAAGAAGAGGCGYGC-----LGSQTSRGRGLGCGAGAAA 145
 DB 176 GKGGDGAAGTSSTAGNSGCGAGSGGKGGADAGAGAGATPGANITAGNGDGD--GAAG 233
 OY 146 AAAAAAGAGGCGGGLGSGQ---GTSCPGYGPQOTSGRGLGCGAGAAAAA 201
 DB 234 AVGISGATGAGDGHGTGAAGNGGTGAGAGSGSIDVGCGGTGCGGNGALGAGAGD 293
 OY 202 AGCAG--GGTGTGCGS-----GTSRGLGCGG--AGAAAAAAGAGAG 244
 DB 294 AGSGSGNGMGITGKGGNAGAGAGAGSNGTGTGANGTGGDGGNGAAGATAGSNGAG 353
 OY 245 GCGYGLGSGTSGRGLGCGG-----AGAAAAAAGAGAG-----GCGY 286
 DB 354 TGSAGNGM--GTGGRGSGGAGGCGIGCGVGGKGGNGADGVEGAGGSGPMTSPGNG 411
 OY 287 GGLSGQTSRGRGLGCGAGAAAAAAGAGAGCGYGLGSGQTSRPG--GGYGPQ 343
 DB 412 GCGGCGGSGGAG--GAAGAGAGGAGNGTANNGCGGAGTGTGAGAAASATNGSGGAG 470
 OY 344 TSGRGLGCGAGAAAAAAGAGAGCGYGLGSGQTSRGRGLGCGG-----A 393
 DB 471 TCGCGSGAGGTGAGAGTCAAGDGGCGCGGAGG--GAGCGAGAGAGTGTGNGNITG 529
 OY 394 GAAAAAAGAG--AGGCGYGLGSGQ---GTSRGLGCGAGAAAAAAGAGAGCGY 450
 DB 530 GTACTAGAGNGAGAGAGAGGCGGTGGGTGGAGAGDGAAGTGGDRTVGGGTVPAGS 589
 OY 451 GGLSGQTSRGRGLGCGAGAAAAAAGAGAGCGYGLG-----SQTSRPGYGP 505
 DB 590 GCGG--GNAGGAGAGG--GGADGSGSGDGDAGTGTGNGNGNRSNGTGTGAGAGNGG 645
 OY 506 QQTSGRGLGCGAGAAA-AAAAAAGAGAGCGYGLGSGQTSRGRGLGCGAGAAAA 563
 DB 646 GAGCAGAGAGGCGGTGTGNGAGAGDAGAGNGNGNGTGTGNGNGNGTGTGAGAGT 705
 OY 564 AAAAAAGAGCGYGL-----GSGT--SGRGLGCGAGAAAAAAGAGAGCGYGLG 618
 DB 706 GSGNGGSGGNGAGNGAGNGSGTGSDDGAGAGNGAGATGTGGDGLTGTGGTGGSG 765
 OY 619 SQTSRGRGLGCGAGAAAAAAGAGAGCGYGLGSGQTSRPGYGPQOTSG 674
 DB 766 --GTGGCGAGNGADMTANMTAGAGDGGGCGGFCGAGAGCGGLTANAGTG 819

DT 01-JUN-1998 (TREMREL. 06, Created)
 DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
 DT 01-OCT-2000 (TREMREL. 15, Last annotation update)
 DE PG8S-FAMILY PROTEIN.
 GN RV0578C OR MT039.16C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID:1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajadaram M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.:
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL021942; CAI17449.1; -
 DR TubercList; RV0578C; -
 DR InterPro: IPR000084; -
 DR InterPro: IPR000209; -
 DR InterPro: IPR002173; -
 DR Pfam: PF00934; PE; 1.
 DR PRODOM: PD001223; -; 1.
 DR PROSITE: PS00583; PKK_KINASES_1; UNKNOWN_2.
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
 SQ SEQUENCE 1306 AA; 105964 MW; 843A30955FFA56B6 CRC64;

Query Match 33.8%; Score 1198.5; DB 2; Length 1306;
 Best local similarity 38.4%; Pred. No. 8.8e-71;
 Matches 323; Conservative 37; Mismatches 293; Indels 189; Gaps 34;

OY 11 GRGSMASGRGLGCGAGAAAAAAG-----AGGCGYGLGSGQTSRGRGLGCGAG 66
 DB 399 GTTSTSGNGDGGKAGADAISSGOTGANGRGCGDGGYVNGAG--GAGGRGAGAGLFG 456
 OY 67 AAAAAAAGAG--AGCGYGC--LGSQTSRGRGLGCGG-----GAAAAA 114
 DB 457 SEADPFGAGGTGAGAGNGGTQAGDGGTGAAGAGGCGGAGAGSIFPNASAPAAAGSP 516
 OY 115 GG-----AGGCGYGL-----GSGTSGRGLGCGG-----GAAA-- 146
 DB 517 GNGNGGPGGAGGAGGAGGAGLALAAAGSNGSAGAGDAGAGAGNGTPTGNGHGAAGALGV 576
 OY 147 --AAAAAAGAGCGYGLGSGQTSG--PGYGC--PGQ--QTSRGLGCGAGAAAAA 199
 DB 577 NGYVGAAGAGHGDPEVGAAGGAGGSGTTPGANGATPTSGNGNGNGRGADATGFGGT 636
 OY 200 AAAG-----AGCGYGLGSGQ-----GTSRGRGLGCGAGAAAA 234
 DB 637 GASGGRGDDGLVNGAGGAGNGSKGLPLGLRLNPLDGTGNGAGAGSGGAGMAGN 696
 OY 235 AAAAAAG--AGGCGYGLGSGQ-----GTSRGRGLGCGG--AGAAAAA 277
 DB 697 GGTGAGGTGGYGTGGSGSDGVNCSAGADCHPGGTGGYGTGKGGDGDGAAPMGV 756
 OY 278 AGAGCGYGLGSGQTSRGRGLGCGAGAAAAAAG--GAGCGYGLGSGQ--GTSGP 335
 DB 757 AGSGPGAGGCG--GTGCGYVNGNGRGIDGADGATAGARQDGGAGAGGKGGGTGTP 814
 OY 336 GGYG-----GQTSRGRGLGSGQ-----AGAAAAA 364
 DB 815 GGAGPAGTTGAGAGNGSGGTGDPDGGGANGANGSVFTNNGTGTGNGNGNGAGPSAG 874
 OY 365 AAGG-----GCGYGLGSGQTSRGRGLGCGAGAAAAAAGAG 407

[illegible]

Db	150	GGSGGNGSGSLGIMNGNGAGGAGGSGGAGAGGNGGMLFGAGGTGCTGAPGAMGCTG	205
Qy	56	GRG-----LCGGCAGAAAAAAGAGAGGCTG-----LGSQTSRGGGLGGGAGGA	105
Db	210	GNGNGALLTGGGGLGAGAGMGGTGGCTGTGTGNGGALLTGGAGVGAGGIGGQGTGA	269
Qy	106	AAAAAAGAGAGAGGCGVCGGLGSQTSGRGGGLGGGAGAAAAAAG-----AGCGYG	161
Db	270	GGAAGAGAGT--GNGAGGCLFMNG--GDCGAGGGGDDGAADAAASAGTGGKGGQGDG	325
Qy	162	GLGSQTSRGPBGYGPCCQOTSGRGGGLGGGAGAAAAAAGAGAGGCGVGLGSQTS---	218
Db	326	GTGGAGAGAVLFGHG---GAGCMGGGQ-----GTGKGAGAGGCTTIV	366
Qy	219	SGRGLGAGGAGAAAAAAGAGAGGCGY--GCLGSQTSRGGGLGGGAGAAAAAAG	276
Db	367	AAGTGGCGCTGGAAGAGCAAGARGALTSGLAGVGAGGTGTGTGGAGDAAAVVFG	426
Qy	277	AAG-----GAGGGGCGGLGSQ-----GTSRGGGLGQ-----	303
Db	427	AMGPGFAGGAGGCGGCTGGAATVTVGAADGCTGKGTGAGAGANDAGSTGNPGKGD	486
Qy	304	-----GAGAAAAAAGAG-----AGCGYGLGSQTSG-----PGYG--	339
Db	487	GGTGGAGAGAGAGTGMGAGAGTGTGGGCGGCTGCGNGGCTGGVNGADNLNPTDFGAGE	546
Qy	340	PGQOTSRRGLGGGAGAAAAAAGAGAGCGYGLGSQTSRGGGLGGGAGAAAAA	399
Db	547	PG-----GAGGAG--GAGGAAGCGPGCTGTGCGNGGNGG--GNGCGNGCGNGAGNNS	598
Qy	400	AAAAAG-----AGCGYGL-----GSGQTSRGGGLGGGAGAAAAAAGAGAGG	449
Db	599	TMAVGGEBGAGGAGGAGGAGAGANGSTAGSQTSGVGEGD-----GAGNGGGG	648
Qy	450	YGLGSQTSG-----RGLGGGAGAAAAAAGAG-----GCGYGLGSQTSG	498
Db	649	KAGTGNNGNTGVSEAGFSSGAGAGNGVGGAAAGANGTGGSGGNGGDDGAGGIGAGGNG	708
Qy	499	PG--GYPGQOTSRRGLGGGAGAAAAAAG-----AGCGYGLGSQTS-----	544
Db	709	IPGTGTAPAGCTGAKGCGDGDGAGAGAGNAGGAGCGGAGCGAGGAGNAAVTPGDGV	768
Qy	545	-----TSGRGLGGGAG-----GAAAAAAGAGAGGCGVGLGSQTS	584
Db	769	GKAPHDAGSGGSGGCGGCGSGSGTGGSGAPTGGGAGCTGGGGAGAGGAGGTAGQST	828
Qy	585	S-----GRGGLGGGAGAAAAAAGAGAGCGGCGGLGSQTSRGGGLGGGAGAAAAA	640
Db	829	TITTPGNGAGNAGDGN---GGMNAGAGNGSGDPFGCNTTSGASGSGGNG--NAGTASGSG	884
Qy	641	AAAAAGAG--GCGYGLGSQTSRGPBGYGPCCQOTSGRIRAPSTS	683
Db	885	AGGTGTGTLGGNGGNGGNGGNGGDDGNGAHGTVGAQF--VPATS	927
RESULT	9		
ID	053215	PRELIMINARY;	PRT; 1660 AA.
AC	053215		
DT	01-JUN-1998	(TREMBLrel. 06, Created)	
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)	
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)	
DE	PGS--FAMILY.		
DE	RV2490C	OR MTV008.46C.	
OS	Mycobacterium tuberculosis.		
OC	Bacteria, Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN=H37R;		
KX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.		

RA Gordon S.V., Eglmeier K., Gas S., Barry C.E. III, Tekela F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osbourne J., Quail M.A., Rajadream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL021246; CAAL6067.1; -.
 DR TubercuList: Rv2490c; -.
 DR InterPro: IPR000084; -.
 DR InterPro: IPR000228; -.
 DR InterPro: IPR002173; -.
 DR Pfam: PF00934; PE: 1.
 DR PRODOM: PD001223; -; 1.
 DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
 DR PROSITE: PS01287; RTC; UNKNOWN_1.
 SO SEQUENCE 1660 AA; 133124 MW; 3A89CE12C0FA945 CRC64;

Query Match 33.3%; Score 1180; DB 2; Length 1660;
 Best Local Similarity 37.6%; Pred. No. 1.7e-69;
 Matches 326; Conservative 38; Mismatches 292; Indels 212; Gaps 35;

QY 2 ASWTGQOMGRSMASRGGLGCGAGAAAAAAGAG-----GCG 45
 DB 514 SSGTPEDEGNGG-ACGACGAGCAHAGDGCAGAGGNGAGCAHFNVLVSDG 572
 QY 46 YGGLSGQTSRGGLGCGAGAAAAAAGAGGCGYGLG- GSGTSGRGLGCGAG 104
 DB 573 NGDGG--GAGGKGGDGGAG-GAGGDAPAGRAGSGVGGDGCAGAGAPNGSGGSDM 629
 QY 105 A---AAAAAAGAGAGCGCGYGLGSG-----TSGRGGLGCGAGAAAA 147
 DB 630 AFKPDGCGAGCGDGPAGGKAGAGATGEGTATGATVHSGNGKCGKNADATVAG 689
 QY 148 AAAAAGCA-----GCGYGLGSGQTSGPGGYGPQQ-----TSGRG 184
 DB 690 ANCGKAGAGNGGLVGDGAGDGGSGAAGANGANGVEDADDTLSGPGESEANGCG 749
 QY 185 GLGSGAGAAAAAAGAGAGCGYGLGSGQTSRGGLGCGAGAAAAAAGAGAG 244
 DB 750 GYGGGAGAGCGDGGAGSSALGSGNGRCDAGG---GAGCAGAGAGAGGSGDGG 805
 QY 245 QCGYGLGSGQTSRGGLGCGAGAAAAAAGAG--GCGYGLGSGQTSRGGLGCG 303
 DB 806 PGKGGAGAGAGAGAGG--GGGKGAAGADSAEAVGAGAGKGGDGYG--GYGDDGPG 862
 QY 304 G--AGAAAAAAGAGAGCGYGLG- GSGQTSGPGGYG-----PGQTSGRG 349
 DB 863 GGAGCAAPAGVCHGVGCGDGLGAGAGNGDGHGSDGDDGDPG--AGGLG 920
 QY 350 LGG--CGAGAAAAAAGAG-----GAGCGYGLGSG-----GTSGR--GG 387
 DB 921 LGGDSGNGTRAAASGVDA SHPGSGGNGGNGAGAAVAGAGAGCGDGNRVRDGG 980
 QY 388 LCGGCGAAGAAAAAAGAGAG-----GGYGLGSG--TSGRGLGCG--GAGAAAA 437
 DB 981 AGNGGDDGAAGANGANGAPGSDALLGDPGNGGCGDAGAGAGAGAGAGGAGG 1040
 QY 438 AAAAAGAGAGCGYGLGSGQTSGR-----GGLGCGAGAAAAAAGAGAGCGY 488
 DB 1041 DGGAGNGAGAGCGYAGAGAGARANGIDTSGTGGAGGCGDGGAGAGVGHGDDG 1100
 QY 489 GG-----LGSQTSGPGGYG-----PGQTS 509
 DB 1101 GGAAPSGTGVSHGTGCGDGLGAGAGVAGAGNGGIGITVGGAGAGAGNGDPP--AG 1158
 QY 510 GRCGLGAG-----GAGAAAAAAGAGAGAGCGYGLGSG 545
 DB 1159 GRCGLGAGDNGTSAANGVDAKHPPLTGDDGCGVGNAGAAAAAGDGGGCGD--GNAG 1216

QY 546 TSGRGLGCGAGAAAAAAGAG--GCGYGG-----LGSQTSGRGLGCGA----- 595
 DB 1217 LTFDGGAGGAGCTAAEALGSDGAGAGKGDADIDGDDGGKGGAGALAGL 1276
 QY 596 -----GAAAAAAGAG--GCGYGLGSGQTSRGGLGCGAGAAA--AAAAAAG 646
 DB 1277 VAGNGAGAGAGAGAGAGFLDGGGNGAGGCGAGAGSGPGGGGCGHGGAGG 1336
 QY 647 AGCGYGLGSGQTSGPGGYGPQQTS 674
 DB 1337 NGGCGTGCGCGAGAGAGAGGWSPSDLKG 1364

RESULT 10
 Q9NHW2 PRELIMINARY; PRT; 1884 AA.
 AC Q9NHW2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
 GN FLAG.
 OS Nephila madagascariensis.
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
 OX NCBI_TaxId=115969;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20156766; PubMed=10688794;
 RA Hayashi C.Y., Lewis R.V.;
 RT "Molecular architecture and evolution of a modular spider silk protein
 gene."
 RL Science 287:1477-1479(2000).
 DR EMBL: AE218623; AAF36091.1; -.
 DR InterPro: IPR000872; -.
 FT NON_TER 1884 1884
 SO SEQUENCE 1884 AA; 148738 MW; 27B6F45339FD20A5 CRC64;

Query Match 32.9%; Score 1167; DB 5; Length 1884;
 Best Local Similarity 40.0%; Pred. No. 1.3e-68;
 Matches 318; Conservative 36; Mismatches 295; Indels 146; Gaps 36;

QY 6 GCGOMGRSMASRGGLGCGAGAAAAAAGAGCGYGLGSG--QTSGRGLGCG 64
 DB 559 GGS--GCGGLGIRSGCGYGP--GSGPGSICPGSGCGGLGPGSGGGLGPGSGPG 615
 QY 65 AGAAAAAAGAGAGCGYGLGSG--QTSGRGLGCGAG-----GAAAAAAG 115
 DB 616 VSGGCGVGPYGPAGSGGCGYAGAGGPGGSGGPGAGGPGGPGYGPAGGPG 675
 QY 116 GAGCGYGLGSGQTSRGGLGCGAGAAAAA--AAAAAGAGCGYG-----GLGSG 168
 DB 676 GAG--GPRGPGPGYGPAGGAGGPGGPGGAGGPGGAGGPGGPGYGPAGGPG 734
 QY 169 SGPYGPQGTSGRGLGCGAG--AAAAAAGAGAGCGY--GGLGSGQTSGRGLG 225
 DB 735 AGPGYGP--GAGPGYGPAGGAGPGGSGPGGIGPGSGPGGIGPGGT--GPGAG 791
 QY 226 GCGAG--AAAAAAGAGAGCGYGLGSG--QTSGRGLGCGAGAAAAAAGAG--G 282
 DB 792 PGAGPGGAGPAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 851
 QY 283 QGGYGLGSG-----GNSGRGLGCGAG--AAAAAAG 316
 DB 852 SGGAGSGGTTTIEDIDTYDANGPTTIEBELTIGAGAGGAGGPGGSGGPGGSG 911
 QY 317 GAGAGCGY--GGLGSGQTSGPYGPQGTSGRGLGCGAG--AAAAAAGAGAGAG 373
 DB 912 GGVPGGSGPGYGP--GSGSGGPGG--GSGPGGPGG--GSGPGGPGGPGGPGG 968
 QY 374 YGGLGSGQTSRGGLGCGAG-----AGAAAAAAGAGAGCGYGLGSG 419

Db	Accession	Length	Score	DB 2	Length
Db	969 VCGEYRPGCGSGPCGAGAGAGSIVPGGPIYRPGSGSGPCGACGPIYRPGAG - GPYGGPGYG	1027			
Oy	420 TSGRGLGSGGAGAAAA -AAAAAGACGGCTGTGCGSCTSGRGLGSGAG - - - - -AA	472			
Db	1028 PGCGGCGGCGPCGAGAPYCPGCGGACGCGYG - - -PGACPGCGPCGACPGCGYCGPG	1083			
Oy	473 AAAAAAGAGAGGGGCGGCGSCTSGPGGYPGQDTSRGRGLGCGGAGAAAAA - - - - -528				
Db	1084 AGSGCYRPGGAGCGCTG - - - -PGPBPCTGTCRGGACPGCTGCTGPGGSAFGAGAPGA	1139			
Oy	529 - - - - -AAGAGAGCGYCGGCGSCTSGRGLGCGGAG - - - - -AAAAAAGAG	572			
Db	1140 GPGGCGGCGSGPGGCGGCGGACPGGAGPCGAGPCGACPGGACPGGACPGGAG	1199			
Oy	573 QCGTG - - - - -GLSGGGS - -GRGLGSGGAG -AAAAAAGACGGCGCT -GS0GTS	624			
Db	1200 PGGAGPEGAPGGGCTGTGCTGCGAGCGAGRGAGRGAGRGAGRGAGRGAGRGAGRGAG	1255			
Oy	625 RCGAGCGAGAAAAA - - - - -AAAAAGAGGCGY -GGLGSGG	659			
Db	1260 AGGCGGAGGSGGTTTIEDLDTITDAGDPTTISELTITSGAGGSGPGGAGPGGAG -PG	1318			
Oy	660 TSGPGCTGPGQDTS	674			
Db	1319 GSGPGGVPGGSGPG	1333			
RESULT	11				
ID	053395	PRELIMINARY;	PRT;	1538	AA.
AC	053395;				
DT	01-JUN-1998 (TRMBLrel). 06, Created)				
DT	01-AUG-1998 (TRMBLrel). 07, Last sequence update)				
DT	01-OCT-2000 (TRMBLrel). 15, Last annotation update)				
DE	PGMS-FAMILY PROTEIN.				
GN	PE_PGRS OR RV3345C OR MTW004.01C-MTV016.45C.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterinae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
LN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37Rv;				
RC	STRAIN=H37Rv;				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
RL	EMBL: AL021841; CAI17117.1; -				
DR	HSSP: P00441; ISOS.				
DR	TubercuList: RV3345C; -				
DR	InterPro: IPR000084; -				
DR	InterPro: IPR002173; -				
DR	Pfam: PF00934; PE; 1.				
DR	ProDom: PD001223; -; 1.				
DR	PROSITE: PS00583; PFR_KINASES.1; UNKNOWN.3.				
SO	SEQUENCE 1538 AA; 129386 MM; 788F0B209587592 CRC64;				

[illegible]


```

Db 756 GGAGAGAGAGGPGGAGGSGCTTIEDLITIDGADPTTISEELTISGAGSGPGGACTG 815
QY 487 GYGGLGS-QGTSRPGYGPQQTSGRGLGCGAGAA-----AAAAAAGGA 533
Db 816 GVGPGGSGPGGVPFGGPGGAGRPYGPBGSSGPGGAGAGTGGA 875
QY 534 -GGGGSQGTSGRGLGCGAGAA-----AAAAAAGGA-----571
Db 876 YGPGGAYPGGSGGPGGEGPGGAGGPGYGPBGAGGPGYGPBGEGGPGY 935
QY 572 -GGGAYG-LSGSGTSGRGLGCGAGAAAAAAGAGAGGCGY-----GLGSG 620
Db 936 GPGYVYGPBGAGGPGGPGGPGGEGPGGAGGPGYGPBGVPGGSGPGYGPBGAG 995
QY 621 GTSRGR-----GLGCGAGAAAAAAGAGAGGCGYGLGS-QGTSRPGCTGPGQQTSG 674
Db 996 GPGGSGPGGPGGSGPG-----GYGPGSGPGGPGGPGGPGGPGGPGGPGG 1048

RESULT 14
Q9GUB5 PRELIMINARY: PRT: 1468 AA.
AC 09GUB5:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HEAVY-CHAIN FIBROIN (FRAGMENT).
GN FIB-H.
OS Galleria mellonella (Max moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Pyraloidea; Pyralidae; Galleriinae; Galleria.
OX NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POSTERIOR SILK GLAND;
RA Zurovec M., Kodrik D., Yang C., Sehmal F.;
RT "Heavy-chain fibroin of Galleria mellonella L.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095239; AAC10393.1; .
FT NON_TER 1468 1468
SQ SEQUENCE 1468 AA; 122705 MW; 2DA59E1181BB3DDE CRC64;

Query Match 32.5%; Score 1152.5; DB 5; Length 1468;
Best Local Similarity 30.1%; Pred. No. 9,4e-68;
Matches 358; Conservative 82; Mismatches 221; Indels 529; Gaps 37;

QY 2 ASMTGGQMGKRSMA-SGRGLGCGA-----GAAAAAAGGA 41
Db 252 ATTSGAGLGAGVAGSAGLGAGAGASAAGSAGAGGAGGAGGSSGSSAASAGSAG 311
QY 42 GQ-----GGYGLGSG 52
Db 312 GEVYILIDRSSAASRAAASGASGVGGLGSLGIPGIGTIGATISASISAGLGVGAA 371
QY 53 GTSRGLGCGA-----GAAAAAAGAGGCG-----81
Db 372 GASGLGAGAGASAAGSAGAGAGAGAGSSGSSAASAGSAGAGEVYIVIDRSSAAS 431
QY 82 -----GGYGLGSGTSGRGLGCGAGAAAA-----AAAAAAGGCGGAGGAGG 131
Db 432 AAAASSGASGLGAGLGAGPYGGLGNGVSSASALGAGLGAGVGTGAGSGLGAGV 491
QY 132 GNG-----GLGCGAGAA-----AAAAAAGAGAGGCGYGL 163
Db 492 AVGPAGAGLGAGAGSSSSAASASASGAPVIVIEDSSAASAAAAAGSASGLG 551
QY 164 GSGGTSGP-GGYGPGQTS-----GRGGLGCGAGAA-----196
Db 552 G-LGAGPGLGIGIPNEVSSASATGAGAGLGGAGSSAASAGAAPAVIYI 610
QY 197 -----AAAAAAGAGGCGYGLGSGTSGRGLGCGG-----228

```

```

Db 611 EDSSAASAAAAAGSAGSLGGLGAMCTLGIGPNVSSASATGASAGSISGLGSGA 670
QY 229 -----AGAA-----AAAAAAG-----241
Db 671 AGSSLSAASAGAAAPVIVIEDSSAASAAAAAGSAGSVGLGALGPLAGIGPICAS 730
QY 242 --GAGGCGYGGISGTSRGLGCGA-----GAAAAAAGG 280
Db 731 SAGASGAGLGAGVAGTAGTGLGIGVAGSTAGSAGAGLGTGAGSSGSSAASAGSAG 790
QY 281 AGQ-----GGYGLGGS 291
Db 791 AGEVIVIDRSSAASAAAAAGSAGSAGCGGLGGLVWGPLGIGPICASASAGAGLGVGA 850
QY 292 QGTSRGLGCGAGAA-----GAGGCGYGGISGTSRGLGCGG 309
Db 851 AGTSGLGILGAGSAGSAGSAGAGLGTGASGSSGSSAASAGTSAGEVYIVIDRSSAA 910
QY 310 AAAAAAGAGGCG-----YGLGSGGTS-----GPGYGPQQTSGRGLGCGG 354
Db 911 SAAAAAGSAGSLGGLGGLGGLGIPYGGIGLNGVSSASALGAGLGVG-----TAGASGLGGL 966
QY 355 AAAAAAAGAGAGGCGGCGYGGISGTSRGLGCGG-----AGAAAAAAGGA 407
Db 967 GTGASAGSAGAGLGAGVAGSAGSAGSAGAGAGEVYIVIDRSSAASAAAAAGSAG 1026
QY 408 GCGG-----YGLGSGGTS-----GRGGLGCGAGAAA-----436
Db 1027 GPGGLGLGVMCPGIGIPNGVSSASATGASAGSTGAGLGGSGAGSSAASAGAAPAP 1086
QY 437 -----AAAAAAGAGGCGYGLG-----SGTSGRGL 464
Db 1087 VIVIEDGSSAASAAAAAGSAGSLGGLGAMCPGIGIPNGVSSASATGASAGSTG-AGL 1145
QY 465 GCGA-----GAA-----AAAAAAGAGGCGYGLGSGTSGP-GGY 502
Db 1146 GGSAGSAGSAGAGAPVIVIEDSSAASAAAAAGSAGSLGGLG-LGAMGPLGIGI 1204
QY 503 GP-----GQTSRGLGCGAGAAA-----AAAAAG 531
Db 1205 GPNVSSASATGASAGSTGAGLGGSGAGSSAASAGAAPVIVIEDGSSAASAAA 1264
QY 532 GAGGCGYGGISGTSRGLGCGGAGAAA-----AAAAAAGAGGCGYGGISGTSRGS- 588
Db 1265 GSGAGSVGLGSLALGPLGIGIPNGVSSASALGAGLGAGAGSGLGTVAGASAGS 1324
QY 589 ---GLGCGAGAA-----AAAAAAGAGGCGYGLGSGG 621
Db 1325 AGAGLGAGVAGSSGLSTASASAGSAGAGEVYIVIDRSSAASAAAAAGSAGSLGGLG 1384
QY 622 TSGRGLGCGAGAAA-----AAAAAAGAGGCGYGLGSGTSGPGYGP 669
Db 1385 LGPYGIGLNGVSSASALGAGLGAGVGTAGASGLGGLGTGASAGSAGAG 1434

RESULT 15
Q06810 PRELIMINARY: PRT: 1329 AA.
AC 006810:
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL 107.4 KDA PROTEIN.
GN PGRS-FAMILY OR RV1450C OR MTCY493.04.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
MD MEDLINE=98295987; PubMed=9634230;

```

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Bredbeck K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornby T., Jajels K., Kiroh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellton S., Squares S., Sultson J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence." ;
 RL Nature 393:537-544(1998).
 DR EMBL: Z95844; CAB09271.1; -.
 DR Tuberculist: RV1450C; -.
 DR InterPro: IPR000084; -.
 DR InterPro: IPR002173; -.
 DR InterPro: IPR003536; -.
 DR Pfam: PF00934; PE: 1.
 DR PRINTS: PR01370; TRNSINTIMINR.
 DR ProDom: PD001223; -. 1.
 DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_2.
 KW Hypothetical protein.
 SQ SEQUENCE 1329 AA; 107410 MW; 777A125F6DBAB234 CRC64;

Query Match 32.5%; Score 1152; DB 2; Length 1329;
 Best Local Similarity 35.8%; Pred. No. 9,4e-68;
 Matches 326; Conservative 35; Mismatches 305; Indels 244; Gaps 35;

QY 7 GQWGRGSMASGRGLGGQA-----GAAAAAAAAAGGA----- 41
 DB 133 GQAGGAGGILMGNGAGSGAPGVGAGAGAGLFTGAGAGAGAGAGAGAGSSGHL 192
 QY 42 GQGGYGLSGSGTSGRGLGGGAGAAAAAGAGGQGYGLSGTSGRGLGGQ 101
 DB 193 GNGVCGAGGQSLIG-GATGAGAGNACLFVGCTGCGPGCGVGV--GTGAGGLGT 249
 QY 102 GAGAAAAAAGAGC-AGCGYGLGSO---GTSRGLGGGAGAAAAA----- 148
 DB 250 LYGAGHGGAGGPPIGVGGHGGVGAAGLLGVGHGGAGHGAEGVAGAAGEDLSPHG 309
 QY 149 -AAAGAGAGGAGYGLG--QGTSGPGYGPQOTSGRGLG-----GQAGAAAA 196
 DB 310 TSGGVGDADGCGTGRGWLAGAGAGAGAGVGTGAGAGAGFSRALIVAGDNGDPGA 369
 QY 197 AAAAAAGAGC--GGYGLSGTSGRGLGGGAGAAAAAGGA-----GQG 246
 DB 370 GAGGTGAGAGTGAHGAAGAGPTSGGNGAGAGAHFSSGKAGNGAGAGAGLVNG 429
 QY 247 GYGLGSGT-----SGRGLGGGAGAAAAAGGA-----GQGYGLG 291
 DB 430 GAGAGAGNAGAPPSGDPNGGGGAGAGAGKGGDGAQAGDAGAGAGKGGNGGA 489
 QY 292 QGTSGRGL-----GQAGAAAAAGGA-----AGCA----- 319
 DB 490 TGAATGLAGADGTGCGKNGAGAGAGGAGGAGGKALAAATHODSMGAGAGAGNGA 549
 QY 320 -GQGYGLSGT-----SGPGYGPQOTSGRGLGGQA----- 355
 DB 550 GGMGDCGNAKFTFNDGCDVGGNGGSGRIGAGAGIGAGSTAGADGARATPTSG 609
 QY 356 -----GAAAAAAAAAGAGGCGYGL-GSQGTSGRGLGGGAGAAAAAGG 406
 DB 610 GNGGTGNGANATVAGAGGAGGAGGNGGLVNGAGAGKGGDGAAGVAGSSPTTAGESGT 669
 QY 407 AGQ-----GGYGLGSO-----GTSRGLGGGAGAAAAAGGA----- 444
 DB 670 SCQNGAGAGAGAGCGRGDRGDDGTGAGAGNAGANATTPGAAGDGGHGGPQAQGN 729
 QY 445 AGQGYGL-----GSQGTSGRGLGGGAGAAAAAGGAGAGGAGY 489
 DB 730 GGQGGPGLAGNLFQNGICQGVGSGGKAGAGLAGDGNAGANGNFAFGDNGGHGNGG 789
 QY 490 --GLSGQTSRPGYGPQO-----TSGRGLGGGAGAAAAAGGAGGAGG 539

DB 790 NPGAGGGGSGAGSTPGAKAHGFTPTSGGDDGGDNGN-----SQVVGNGGDCGNG 844
 QY 540 GL-GSQGTSGRGLGGGAGAAAAA-----AAAGAGGCGYGLGSG 583
 DB 845 GNGSAGTGGNGRGCDCAFEGMSANATNPGENPNGNPGNGAGAGAGAGLNG-GNGC 903
 QY 584 TSGRGLGGGAGAAAAAAGAGCQ---GGYGLG-----SQGTSGR 625
 DB 904 AGNGGLGEGFGNAGANGVAVGAPQPGAGGHGGAGGNGAGNGCGGVSDGAGA 963
 QY 626 GGLGGC-----AGAAAAAGAGAGGAGYGL-LGSGTSGPGYGPQOTSGIR 676
 DB 964 GAGGDCGAPDGAANGAGAGAFAGGCGRGDGGNAGAGAGTGTGTAGKAG-- 1021
 QY 677 IRAPSTFEH 686
 DB 1022 ---PAGSTLH 1028

Search completed: July 3, 2001, 15:05:35
 Job time: 600 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:11 ; Search time 86.32 Seconds
(without alignments)
370.823 Million cell updates/sec

Title: US-09-490-291-8
Perfect score: 2700
Sequence: 1 MASMTGGCGMGRIRIRGYCG.....GLSGCDVLQALLGHHHHH 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID58/gcgdata/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/AA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length DB	ID	Description
1	2319	85.9	646	18	AAW27178
2	2196	81.3	718	19	AAW53346
3	2196	81.3	718	21	AAV59070
4	2182	80.8	718	12	AAV4308
5	1999	74.0	651	20	AAV40097
6	1883.5	69.8	606	16	AAV39035
7	1883.5	69.8	606	20	AAV40102
8	1879.5	69.6	606	20	AAV40101
9	1871.5	69.3	604	16	AAV39057
10	1858.5	68.8	606	16	AAV39053
11	1858.5	68.8	606	20	AAV40100

12	1344	49.8	831	16	AAV80168	PMISS1 Misp spider
13	1234	45.7	615	20	AAV40099	Spider silk proteol
14	1144	42.4	595	12	AAV4309	N.clavipes draglin
15	1144	42.4	595	19	AAV53347	Nephila clavipes s
16	1144	42.4	595	21	AAV59071	N. clavipes spider
17	973.5	36.1	465	22	AAV70188	Pepide dendrimer
18	972	36.0	1177	9	AAV80940	SIPRII protein com
19	972	36.0	1177	11	AAV80307	SLP III (silk-fibr
20	972	36.0	1177	17	AAV85105	Silk-like protein
21	972	36.0	1177	18	AAV26342	Silk-like protein
22	972	36.0	1177	21	AAV78277	SIPRII amino acid
23	972	36.0	1177	22	AAV72725	Repetitive protein
24	972	36.0	1177	22	AAV63995	SLPRII protein seq
25	972	36.0	1178	14	AAV41007	Silk-like sequenc
26	972	36.0	1178	19	AAV53518	Amino acid sequenc
27	970	35.9	714	16	AAV39059	Spider dragline va
28	970	35.9	714	20	AAV40103	Polymer of an anal
29	962.5	35.6	1023	19	AAV53524	Amino acid sequenc
30	962.5	35.6	1059	14	AAV41013	SLP4 multimeric pr
31	962.5	35.6	1059	18	AAV26348	SLP4 synthetic pro
32	962.5	35.6	1101	21	AAV78283	SEPL4 protein comp
33	955.5	35.4	1059	9	AAV82962	Adhesion protein.
34	946.5	35.1	980	16	AAV81318	Silk-like protein
35	929	34.4	1332	17	AAV85109	Repetitive protein
36	929	34.4	1332	22	AAV72729	SLP-C protein sequ
37	929	34.4	1332	22	AAV63999	Fibronectin cell b
38	922.5	34.2	1038	17	AAV85107	Repetitive protein
39	922.5	34.2	1038	22	AAV72727	FCB-SLP protein fr
40	922.5	34.2	1038	22	AAV63997	Amino acid sequenc
41	921.5	34.1	766	18	AAV26349	FCB-SLPRII (57 kDa
42	921.5	34.1	766	21	AAV78285	FCB-SLPRII amino a
43	921.5	34.1	979	21	AAV78286	FCB-SLPRII amino a
44	921.5	34.1	979	21	AAV78286	FCB-SLPRII amino a
45	921.5	34.1	1050	18	AAV26350	FCB-SLPRII (72 kDa

ALIGNMENTS

RESULT 1	AAW27178	standard; Protein: 646 AA.
AC	AAW27178;	
DT	09-DEC-1997	(first entry)
DE	Nephila clavipes spider silk protein.	
KW	High strength film; fibre; woven article; parachutes; sails;	
KW	absorber; body armour; heavy metal; biological weapon; chemical;	
KW	flavour; fragrance; Nephila clavipes.	
OS	Nephila clavipes.	
PN	W09708315-A1.	
PD	06-MAR-1997.	
PE	22-AUG-1996;	96WO-US13767.
PR	22-AUG-1995;	95US-0517694.
PA	(BASE/) BASEL R M.	
PA	(ELIO/) ELION G R.	
PI	Basel RM, Elion GR;	
DR	WPI; 1997-179272/16.	
DR	N-PSDB; AAT85356.	
PT	New opt. multimerised DNA sequences encoding spider silk protein -	
PT	contg. both repetitive and non-repetitive sequences, useful for	

	Query Match	81.3%	Score 2196:	DB 21:	Length 718:	
	Best Local Similarity	70.3%:	Pred. No. 6.4e-144:			
	Matches 479:	Conservative	4:	Mismatches 10:	Indels 188:	Gaps 15:
OY	6 GGOQGRIRIRRYGGLGCGAGCAGAGAAAAAAGAGAGCGYGGLGSOGAGRGCAGAAA	65				
	:					
Db	22 gqgagagq---ggysglsgygaagagagaataaaagagaggyyglsysgaagrsgagaaa	78				
OY	66 AAAGAGCGCGYCTGLSSQAGRCGLCGQCAG--AAAGAAGVGGCLG---	119				
Db	79 aaagagaggyyyglsysgaagrsglsysgaagaaaaaagagaggyyglsynagaar-gygqa	138				
OY	120 AAAAAGACAGCGYCTGLSSQAGRCSSCGCAGAAAAAAG--	166				
Db	139 aaaaagagaggyyglsysgaagrsglsysgaagaaaaaagagaggyyglsysgaaggyyg	198				
OY	167 GLCSGACRRCGLCGCGAGCA-----	185				
Db	199 glgsagagrrgslsgygaagaaaaaagagaggyylsgygaagagagaataaagagaggyyg	258				
OY	186 -----AAAAAAGAGCGCYGGLGSGAGCGCGYCTGLSSQAGRCGLGCGAGCA	232				
Db	259 gsgaggrsgagagaaaaaagagaggyyglsysgaaggyyglsysgaagrsglsysgaag-	317				
OY	233 AAAAGACAGCGLG---GCGACAAAAAGACAGCGYCTGLSSQAGR--GCGACA-AA	284				
Db	318 aaaagagaggyyglsysgaagagaaaaaagagaggyyglsysgaagrsglsysgaavaaa	377				
OY	285 AAGGAGCGCGYCT---GAGCGC-----GYGLGSGOAGRCGLGCGAGCA	324				
Db	378 aagagaggyyglsysgaagrsgagaaaaaagagagrrgysglsgygaagrsglsysgaaga	437				
OY	335 AA-----AAAAAGAGAQ-----	336				
Db	438 aaaaaagagaggyyglsynagaagrsgagaaaaaagagaggyyglsysgaagrsgagaaa	497				
OY	337 -----GGLGCGCAAAAAAGACAGCGCLGCGAG	365				
Db	498 aaavgagagrlsgygaaggyyglsysgaagrsglsysgaagaaaaaagagaggyyglsysga	557				
OY	366 OGAGAAAAAAGAGVRGCGYCTGLSSQAGRCGCGAGAAAAAGAGCGCGTGLGCGCV	425				
Db	558 ggaq---aaaaaagvrtyygyslsgygaagrsgygaagaaaaaagagaggyyglsysga	613				
OY	426 GAGGLGCGAGAAAAAVGAGCGCYGCVGSCASAAASRLSSPQASSRVSAVNMLVASG	485				
Db	614 grgygsgygaagagaggyygyslgvsgasasaasrastrspqssrsvsaavnlvassg	673				
OY	486 PTNSALSTTSINVSQIGAS	506				
Db	674 ptnsaalstslsnvvslgas	694				
RESULT	4					
AARI4308	AAARI4308 standard; Protein: 718 AA.					
XX	AAARI4308;					
XX	AC					
DT	15-JAN-1992 (first entry)					
XX	XX					
DE	N.clavipes dragline silk protein-1.					
XX	XX					
KW	protein superfibre; major ampullate silk; orb web spider.					
OS	Nephilila clavipes.					
XX	XX					
PN	EPA52925-A.					
XX	XX					
DD	23-OCT-1991.					
XX	XX					
PF	18-APR-1991; 91EP-0106217.					
XX	XX					

[illegible]

RESULT	6
AA899055	
ID	AA899055 standard; Protein; 606 AA.
XX	
AC	AA899055;
XX	
DT	17-JAN-1997 (first entry)
XX	
DE	Spider dragline variant, DP-1B.9 polymer.
XX	
KW	Spider; dragline protein; variant; monomer; polymer;
KW	fiber forming region; Spidroin 1; Nephila clavipes; DP1; mimic;
KW	DP-1A analogue; fiber; high tensile strength; elasticity; clothing;
XX	rope; surgical suture; implant; reinforcement; film; coating.
OS	Synthetic.
XX	
PN	WO9429450-A2.
XX	
PD	22-DEC-1994.
XX	
PF	15-JUN-1994; 94WO-US06689.
XX	
PR	15-JUN-1993; 93US-0077600.
XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.
PI	
XX	Fahnestock SR;
XX	
DR	WPI; 1995-036479/05.
PT	New synthetic variants of spider dragline protein - for making
PT	fibres useful as clothing, surgical silk, plastic reinforcement
PT	etc., also related DNA, vectors and transformed cells
XX	
XX	Disclosure; Page 88-90; 168pp; English.
CC	
XX	This sequence represents a synthetic spider dragline variant polymer,
CC	DP-1B.9. The sequence of the DP-1B.9 monomer is given in AA899034.
CC	The polypeptide monomer is a variant based on a consensus sequence
CC	derived from the fiber forming regions of spider dragline protein.

CC esp. the natural protein 1 (Spidroin 1) from *Nephila clavipes*. The
CC DNA sequence encoding the monomer may be used in the recombinant
CC production of the variant protein in a recombinant host, e.g. *E. coli*
CC or *Bacillus subtilis*. Synthetic analogues of DPI were designed to mimic
CC the repeating consensus sequence of the natural protein and the pattern
CC of variation among individual repeats. This monomer exhibits all of the
CC regularities of (1)-(5) below. In addition, it exhibits a regularity o
CC the natural sequence which is not shared by DP-1A, namely that a repeat
CC in which both GYG and GRG are deleted is generally preceded by a repeat
CC lacking the entire poly-alanine repeat, with one intervening repeat.
CC The sequence of DP-1B matches the natural sequence more closely over
CC a more extended segment than does DP-1A. The individual repeats differ
CC from the consensus sequence given in AAM06201 according to the pattern:
CC (1) the poly-alanine sequence varies in length from 0-7
CC residues; (2) when the entire poly-alanine sequence is deleted,
CC so also is the surrounding sequence encompassing AGRGGGCGAGAGCG;
CC (3) aside from the poly-alanine sequence, deletions usually
CC encompass integral multiples of three consecutive residues;
CC (4) deletion of GYG is generally accompanied by deletion of GRG
CC in the same sequence; and
CC (5) a repeat in which the entire poly-alanine sequence is
CC deleted is generally preceded by a repeat containing six alanine
CC residues.
CC The proteins may be used to produce fibres of high tensile strength and
CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC for implants, plastic reinforcements, films, coatings, etc.

Query Match	69.8%	Score 1883.5	DB 16	Length 606
Best Local Similarity	69.6%	Pred. No. 1.7e-122		
Matches 412	Conservative 3	Mismatches 26	Indels 151	Gaps 16
QY 6 GGOOMGRIRI-----RCGGLGCGGAGGAGCAAAAAAGCAAGCGGCGGLGSGGAGRCGCG 60				
Db 21 ggggagaaaaaaagagagggglgsggagagag -aaaaaaggggggylgsggagrrggg 79				
QY 61 AGAAAAAG-----GACGGCGGLGSGGAGRGGLGCGGAG--AAAAAGCGQ 105				
Db 80 agaaaaaggaagggggylgsggaaggggylgsggagrrgglgsggaagaaaaagagag 139				
QY 106 GGLGCGGAGCGGAGAAAAAGCAGCGCGGGLGSGGAGRCGSGCGGAGAAAAAG-- 158				
Db 140 gglgsggagggagaaaaaagagggggylgsggagrr -gggagagaaaaaagaggggyl 196				
QY 159 -----GAGCGGCGGLGSGGAGRGGLGCGGAG----- 185				
Db 157 gglgsggagggggggylgsggagrrglgsggagaaaaaagggggylgsggaggaagaa 256				
QY 186 -----AAAAAGCAGCGGCGGLGCGGAGCGCGGGLGSGQ 218				
Db 257 aaaagagggggylgsggagrrggagagaaaaaagggggggylgsggagggggylgsg 316				
QY 219 GAGRGGLGCGGAG--AAAAAGCAGCGGGLG-----CGCAGAAAAAGCAGCGCGGLGSGGA 272				
Db 317 gaggrrlggggagaaaaaagaggggylgsggagggagagaaaaaagggggylgsgga 376				
QY 273 GRCGCGAGAAAAAGCAGCGGCG--CGACGGCGGLGSGGAGRGGLGCGGAGAAAAA 329				
Db 377 grrggagagaaaaaagagggggylgsggagggggylgsggagrrglgsggag -aaaa 435				
QY 330 AAGGAGCGGGLG-----CGCAGAAAAAGCAGCGGGLG--CGAGCGCAGAAAAAGAGV 382				
Db 436 aagggggggylgsggagggagagaaaaaagggggylgsggagrr -ggagagaaaaaagaa 494				
QY 383 RCGGCGGLGSGGAGR-----GAGCGCGGLGSGGAGRGGLGCGGAG--GCG 401				
Db 485 gggggggylgsggagggggylgsggagrrglgsggagaaaaaagggggylgsggaggg 554				
QY 402 AGAAAAAGGAGCGGGLGCGGAGGAGGLGCGGAGAAAAAGCAGCGCGGCGVGS 453				
Db 555 agaaaaaagggagggggylgsggagrrggga -gaaaaaagggggggggylgsg 605				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:41 ; Search time 48.42 Seconds

(Without alignments)
219,669 Million cell updates/sec

Title: US-09-490-291-8

Perfect score: 2700
Sequence: 1 MASMTGCGQGMKIRIRGCG.....GLSCDVLIALGLHHHHH 528Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/3A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/9A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/9B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	84.2	747	4	US-09-034-177-3 Sequence 3, Appl1
2	2196	81.3	718	1	US-08-425-069-2 Sequence 2, Appl1
3	2196	81.3	718	2	US-08-317-844B-2 Sequence 2, Appl1
4	1343.5	49.8	832	1	US-08-209-747-2 Sequence 2, Appl1
5	1343.5	49.8	832	1	US-08-458-298-2 Sequence 2, Appl1
6	1144	42.4	595	1	US-08-425-069-4 Sequence 4, Appl1
7	1144	42.4	595	2	US-08-317-844B-4 Sequence 3, Appl1
8	972	36.0	1177	1	US-07-609-716-31 Sequence 31, Appl1
9	972	36.0	1177	1	US-08-175-155-29 Sequence 29, Appl1
10	972	36.0	1177	1	US-08-477-509B-64 Sequence 64, Appl1
11	972	36.0	1177	2	US-08-707-237A-35 Sequence 35, Appl1
12	972	36.0	1177	3	US-08-482-085B-64 Sequence 64, Appl1
13	972	36.0	1177	4	US-08-475-411A-31 Sequence 31, Appl1
14	972	36.0	1177	4	US-08-478-029A-31 Sequence 31, Appl1
15	962.5	35.6	1059	2	US-08-175-155-48 Sequence 48, Appl1
16	962.5	35.6	1059	2	US-08-707-237A-54 Sequence 54, Appl1
17	962.5	35.6	1101	1	US-08-477-509B-83 Sequence 83, Appl1
18	962.5	35.6	1101	3	US-08-482-085B-83 Sequence 83, Appl1
19	929	34.4	1332	1	US-07-609-716-41 Sequence 41, Appl1
20	929	34.4	1332	4	US-08-475-411A-41 Sequence 41, Appl1
21	929	34.4	1332	4	US-08-478-029A-41 Sequence 41, Appl1
22	922.5	34.2	1038	1	US-07-609-716-36 Sequence 36, Appl1
23	922.5	34.2	1038	4	US-08-475-411A-36 Sequence 36, Appl1
24	922.5	34.2	1038	4	US-08-478-029A-36 Sequence 36, Appl1
25	921.5	34.1	766	1	US-08-175-155-53 Sequence 53, Appl1
26	921.5	34.1	766	1	US-08-477-509B-88 Sequence 88, Appl1
27	921.5	34.1	766	2	US-08-707-237A-61 Sequence 61, Appl1

28	921.5	34.1	766	3	US-08-482-085B-88 Sequence 88, Appl1
29	921.5	34.1	979	1	US-08-477-509B-89 Sequence 89, Appl1
30	921.5	34.1	979	3	US-08-482-085B-89 Sequence 89, Appl1
31	921.5	34.1	1050	1	US-08-175-155-54 Sequence 54, Appl1
32	921	34.1	745	2	US-09-010-928B-28 Sequence 28, Appl1
33	921	34.1	870	2	US-09-010-928B-2 Sequence 2, Appl1
34	911.5	33.8	1018	1	US-08-089-862-11 Sequence 11, Appl1
35	911.5	33.8	1018	1	US-08-587-333-18 Sequence 18, Appl1
36	911.5	33.8	1018	5	PCT-US94-07776-16 Sequence 16, Appl1
37	895.5	33.2	1011	1	US-08-477-509B-94 Sequence 94, Appl1
38	895.5	33.2	1011	3	US-08-482-085B-94 Sequence 94, Appl1
39	895.5	33.2	1170	1	US-08-175-155-59 Sequence 59, Appl1
40	895.5	33.2	1170	2	US-08-707-237A-66 Sequence 66, Appl1
41	890.5	33.0	738	3	US-08-864-038A-3 Sequence 3, Appl1
42	889.5	32.9	784	1	US-07-609-716-48 Sequence 48, Appl1
43	889.5	32.9	784	4	US-08-475-411A-48 Sequence 48, Appl1
44	889.5	32.9	784	4	US-08-478-029A-48 Sequence 48, Appl1
45	882	32.7	2100	1	US-08-477-509B-80 Sequence 80, Appl1

ALIGNMENTS

RESULT 1
US-09-034-177-3
Sequence 3, Application US/09034177
Patent No. 6127146
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN FIBROUS PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,177
FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0486 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 amino acids
TYPE: amino acid
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: GI 1174414
US-09-034-177-3
Query Match 84.2% Score 2273; DB 4; Length 747;


```

: TISSUE TYPE: minor ampullate gland
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..309
US-08-209-747-2

```

Query Match	49.88;	Score 1343.5;	DB 1;	Length 832;
Best Local Similarity	47.98;	Pred. No. 1e-91;		
Matches 343;	Conservative 19;	Mismatches 143;	Indels 211;	Gaps 25;

QY 2 ASMTGGGQQMGRIIRTRIGYGGLGGGAGCGAGCAAAAAAAGGAGGCGYG-----G 48

Db 93 AAGAAGCYGRG---AGYGVGGCGGYGAGAGAGCAAAAAAGACAGAGAGCGYGRGAGAGAAAAG 149

OY 49 LGSOGAGRGGQ----GAGAAAAAGAGCGGYGTLG-----SQGARGSLGGQ-- 92

Db 150 AGAGGAGYGGGGYGAGAGCAAAAAAGAGAGGAGCYRGAGAGAGAAAGAGAGGGGGG 209

OY 93 -----GACAAAAACGVGGCGGLGGGAGAGAGAAAAAGAGAGCGGYGTLG----- 136

Db 210 YGACAGAGCAAAAAAGAGSGGCMGTGRCAGCAGACGAACAAGCACAGSYGGCGYTGAGAGAGAAA 265

OY 137 -----SQGARGGSSG--GGGAGAAAAAGAGCGGGYGTLG----- 169

Db 270 AAXXXXXXXXXXXXXXKAGAGNAYGGCGGCGYGAGAGCAAAAAAGAGAGAGYGRGAGAG 329

OY 170 -----SQCAGRGLGGQ-----GACAAAAAAG----- 132

Db 330 AGAAAGAACAGAGAGYGGGDSYGCAGAGAAAAAGAGAGAGYGRGAGAGAGAAAGAGACAA 389

OY 193 GAGCGYGGTLCGGAGCGGYGGLTSQGAGRGCLTCCGAGCAAAAAAG--GAGCGGLGGQ-- 247

Db 390 GAGAGYGGCGGYGAGAGA-GAAAAAGAGAGAGCGYGRGAGAGCAAAAAAGAGAGCGGCG 448

OY 248 ----GACAAAAAAGA-GCGGYG-----GLSSQAGRGGQ-----GAGA 281

Db 449 YGAGAGACAAAAATTCAGCAGCGYGRGAGAGAGAAAAAGAGAGYGGCGGYGAGAGAGA 508

OY 282 AAAAAAGAGCGGYG-----CGAGCGGYGGLSSQAG----- 313

Db 509 AAAAGAGAGGAGYGGAGAGAGAAAGAGAGAAAAAGAGAGYGGCGGYGAGARAGAAAAA 568

OY 314 -----RGTLGGCGAGAAAAAAGAGCGGGLGGQ-----GACAAAAAAGAGCGG 357

Db 569 GAGAGAGYGRGCRAG-MAAGAGAAAAAGAGAGAGYGGCGGYGAGAGCAAAAAAGAGSG 627

OY 358 GLCGCGAGCGAAAAAAGAGYRGGYGLCSQAGRGCGCAGAAAAAAGAGCGGCT 417

Db 628 GAGGYGRGAGAAAAAGAAAAAGAGAGAGYGGCGGYGCH-----GAGAAAAAGAGAGRGY 682

OY 418 GGLGGCGYAGAGLGGQ-----GAGAAAAAGAGCGGYGCVGSGA-----SAASAAARS 465

Db 683 ---GRGAGAGYGGGGYGAGAGAGAAAAAGAGAGGGIDKEIKCWSRCRYVASTTSSL 738

OY 466 SSPQASSSVSSAVSNLYASGPTNSAALSTISNVYSOIGASNPLGSCDVLIOALL 521

Db 739 SSAEKSNTISSAATIVSGYLNTAALPVSISDLFAQYAGASSPYTRQRS-LIQYLL 793

RESULT 5
US-08-458-298-2
Sequence 2, Application US/08458298
Patent No. 5756577
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Colgin, Mark
TITLE OF INVENTION: CDNs Encoding Minor Ampullate Spider
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747

CITY: Falls Church
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22040-3487
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, V
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/458,298
 FILING DATE: 02-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/209,747
 FILING DATE: 14-MAR-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1447-104B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-205-8000
 TELEFAX: 703-205-8050
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 832 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 FRAGMENT TYPE: internal
 ORIGINAL SOURCE:
 ORGANISM: N. clavipes
 TISSUE TYPE: minor ampullate gland
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..309
 US-08-458-298-2

	Query Match	Similarity	49.8%;	Score 1343.5;	DB 1;	Length 832;	
	Best Local Similarity		47.9%;	Pred. No. 1e-91;			
	Matches 343;	Conservative 19;	Mismatches 143;	Indels 211;	Gaps	25;	
OY	2	ASMTGGQOMGRIRINGYGGLGSGQGAGAGAAAAAAGAGCGGYG-----G	48				
Dd	93	AAGAGCYGRG---ACGYGGGGYGAGAGAGAAAAGAGAGGAGCGRAGAGAGAAG	149				
OY	49	LGSQAGRGQ-----GAGAAAAAGCAQCQGYGGLG-----SOGARBGILGQ--	92				
Dd	150	AGAGAGAGYGGQGCYGAAGAGAGAAAAAGAGAGGAGGYGRRAGACAGAAAGAGAGTGGQGG	209				
OY	93	-----GAAAAAAAGGVGGGGLGGQAGAGAGAAAAAAGAGCGGYGGLG-----	136				
Dd	210	YGAGAGAGAAAAAAGAGSGAGGYGRRAGAGAGAAAAGAGAGAGTGGCQYGAAGAGAA	269				
OY	137	-----SOGARGGSS--GQAGAAAAAAGCAQCQGYGGLG-----	169				
Dd	270	AAAXXXXXXXXXXXXXXGAGAGAGYGGGCGYAGAGAGAAAAAAGAGAGGAGGYGRRAGAG	329				
OY	170	-----SOGAGRGLGQ-----GAAAAAAAAG--	192				
Dd	330	AGAAAGAGAGAGAGYGGQBSYGAAGATAAAAAAGAGAGGAGGYGRRAGACAGAAAGAGAA	389				
OY	193	GAGQCGYGLGGQAGAGCGGYGGLSGQAGRGGLGQQAGAAAAAG----GAGCGGLGQ--	247				
Dd	390	GAGAGCGYGGQCGYGAAGAG--GAAAAAGAGAGAGGAGGYGRRAGAGAGAAAGAGYGGQGG	448				
OY	248	-----GAGAAAAAAGGA--GGGCGY-----GLGSGQAGCGCG-----GACA	281				
Dd	449	YGAGAGAGAAAAATGAGAGAGGYGRRAGACAGAAACAGACTGTGAGYGGQGGYGAAGAGAA	508				

OY	282	AAAAAGGAGCGGCGT-----GGAGCGGTGGTGGGAG-----	31.3%
Db	509	AAAAGAGGAGCTRCGACAGCAAAAGACAGAAAGACGAGTGGCGTCGACARGAALAA	56.6%
OY	314	-----RGLGCGCAGCAAAAAAAAAAGCGGTLGGQ-----GAGAAAAAAGAGCG	35.7%
Db	569	CAGAGCAAGTCRGCRAG- AAGAGAGGAAAAGAGAGAGTGCGGCGTAGAGAGCAAAAAAGAGSG	62.2%
OY	358	GLGGGAGGOGAGAAAAAAAAGGVRGGCGTGCAGSAGRRGGGAGAAAAAGAGAGCGGT	41.7%
Db	628	GAGCTGCGAGCAAGAAAAGAAAAGACAGACATGCGCGTYRA-----GAGAAAAAAGAGAGAGCT	68.8%
OY	418	GGTGGCGVAGCTLGGQ-----GAGAAGAAVAGCGGTGGVGSA-----SASAAARSRL	46.1%
Db	683	----GRGAGAGGTGCGCGCGTAGACAGAGAAAAAGAGAGCTGAKETICWSCRCHTVASTSRLL	73.8%
OY	466	SSPQASVVSSAVNVLVAASPTNSAALLSTSNVNVISIGANSPELSCDYLIOALL	52.1%
Db	739	SAAASSTRISSAASTLVSGCTLNLTALPSVTSDLFADYGASSPVITRRRS-LIQVLL	79.3%

```

1      RESULT      6
2      US-08-425-069-4
3      : Sequence 4: Application US/08425069
4      : Patent No. 5728810
5      : GENERAL INFORMATION:
6      : APPLICANT: Lewis, Randolph V.
7      : APPLICANT: Xu, Ming
8      : APPLICANT: Himman, Michael B.
9      : TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
10     : TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
11     : TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
12     : NUMBER OF SEQUENCES: 69
13     : CORRESPONDENCE ADDRESS:
14     : ADDRESSEE: Birch, Stewart, Kolasch & Birch
15     : STREET: 301 No. 5728810th Washington Street
16     : CITY: Falls Church
17     : STATE: Virginia
18     : COUNTRY: U.S.A..
19     : ZIP: 22046
20     : COMPUTER READABLE FORM:
21     : MEDIUM TYPE: Floppy disk
22     : COMPUTER: IBM PC compatible
23     : OPERATING SYSTEM: PC-DOS/MS-DOS
24     : SOFTWARE: PatentIn Release #1.0, Version #1.25
25     : CURRENT APPLICATION DATA:
26     : APPLICATION NUMBER: US/08/425,069
27     : FILING DATE: 19-Apr-1995
28     : CLASSIFICATION: 435
29     : ATTORNEY/AGENT INFORMATION:
30     : NAME: Murphy Jr., Gerald M
31     : REGISTRATION NUMBER: 28,977
32     : REFERENCE/DOCKET NUMBER: 1447-106P
33     : TELECOMMUNICATION INFORMATION:
34     : TELEPHONE: (703) 205-8000
35     : TELEFAX: (703) 205-8050
36     : TELEX:
37     : INFORMATION FOR SEQ ID NO: 4:
38     : SEQUENCE CHARACTERISTICS:
39     : LENGTH: 595 amino acids
40     : TYPE: amino acid
41     : TOPOLOGY: linear
42     : MOLECULE TYPE: protein
43     : US-08-425-069-4

```

	Query Match	43.4%	Score 1144;	DB 1	Length 595;
	Best Local Similarity	51.1%;	Pred. No.	3,4e-77;	
	Matches 313;	Conservative	25;	Mismatches 139;	Indels 136; Gaps 30.
Oy	17 GYG-----GLGGGAGQ----	GAG-AAAAAAAAAAGACGGGTG-----	GLASGCGAGR--	56	
Db	3 GYGPQGQSPGGTGPQQGSPSGSAAAAAAMAPGCTGGPQGGPGCGTGGGRG	62			

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

REFERENCE/DOCKET NUMBER: 1447-105P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 241-1300
 TELEFAX: (703) 241-2848
 TELEX: 248345
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 595 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-317-844b-4

Query Match 42.4%; Score 1144; DB 2; Length 595;
 Best Local Similarity 51.1%; Pred. No. 3.4e-77;
 Matches 313; Conservative 25; Mismatches 139; Indels 136; Gaps 30;

```

OY 17 GYG----GLGGGAGG----GAG-AAAAAAAAAGAGGGG-----GLSGGAGR-- 56
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 GYGCGGCGPGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62
OY 57 -----GGGAGAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 PGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122
OY 104 -GGGGLGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 SGCGGPGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 171
OY 158 GGAGG---GGYG---GLSGGAGRGGGLGGGAGGAGGAGGAGGAGGAGGAGGAG 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 SGPGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 231
OY 211 GYGGLGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 G-----GGGCGGGLSGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 285
OY 262 GGYG----GLSGGAG-RRGGGAG-AAAAAGAGAGGCGGCGGCGGCGGCGGCGG 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 GYGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 342
OY 316 GLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 GPGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
OY 359 LGGGGAGG---GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 PGGTAPEGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 459
OY 408 AAGGAGGCGG---GLGGGAG---GAGGLG-----GGGAGGAGGAGGAGGAGG 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 AAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 519
OY 448 YGGVSGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 YG----PGSQAASAAARLSPDSGARVASAVSNLVSPTSSAALSYISNAVSIGASN 575
OY 508 PGLSGCDVLIAL 520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 PGLSGCDVLIAL 588
  
```

RESULT 8
 US-07-609-716-31
 Sequence 31, Application US/07609716
 Patent No. 5514581
 GENERAL INFORMATION:
 APPLICANT: Ferrari, Franco A.
 APPLICANT: Cappello, Joseph
 TITLE OF INVENTION: Functional Recombinantly Prepared
 TITLE OF INVENTION: Synthetic Protein Polymer
 NUMBER OF SEQUENCES: 118

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: CA
 COUNTRY: US
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/609,716
 FILING DATE: 06-NOV-1990
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rowland, Bertram I
 REGISTRATION NUMBER: 20015
 REFERENCE/DOCKET NUMBER: A-55186-3/BIR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-781-1989
 TELEFAX: 415-398-3249
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1177 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-609-716-31

Query Match 36.0%; Score 972; DB 1; Length 1177;
 Best Local Similarity 44.1%; Pred. No. 3e-64;
 Matches 243; Conservative 30; Mismatches 224; Indels 54; Gaps 17;

```

OY 17 GYGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 GAGAGSGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 692
OY 71 AGGGGGLGGS--QGAGRGGLGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 AGAGSGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 752
OY 130 GYGGLGGS--QGAGRGGSGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 GAGSGAGAGSGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812
OY 185 AAAAAAG---GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 AGSGAGAGSGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 872
OY 238 -----GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 GSGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 932
OY 285 AAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 933 SGAGAGSGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 992
OY 345 GAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 GAGAGSGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1052
OY 399 ---GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1053 AGAGSGAGAGSGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1112
OY 456 SAASAAARLSPDASRVSSAVSNLVSPTNSAALSTTISNVVSQIGASNPGLSGCDV 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1113 GAGSGAG-----AGSGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1159
  
```


OY 516 LIOALLGHHH 526
Db 1160 RYOLSGRHYH 1170

RESULT 9

US-08-175-155-29
; Sequence 29, Application US/08175155
; Patent No. 5641648
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: Methods for Preparing Synthetic
; TITLE OF INVENTION: Repetitive DNA
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albilton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,155
; FILING DATE: 29-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-5/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-175-155-29

Query Match 36.0%; Score 972; DB 1; Length 1177;
Best Local Similarity 44.1%; Pred. No. 3e-64;
Matches 243; Conservative 30; Mismatches 224; Indels 54; Gaps 17;

OY 17 GYGLGCGGAGGAGAAAAAG-----GAGGCGYGLGCGAGRG-GGAGAGAAAAAG 70
Db 633 GAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 692
OY 71 AGGCGYGLG-OGAGRGCGGAGAAAAAGGCGGAGCGAGGAGAAAAAGAGQ 129
Db 693 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 752
OY 130 GYGLGCGGAGGAGAAAAAG--GAGGCGYGLG-OGAGRGCGGAG 184
Db 753 GAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 812
OY 185 AAAAAAG---GAGGCGYGLG-GGAGGCGYGLG-OGAGRGCGGAGAAAAAG- 237
Db 813 AGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 872
OY 238 ----GAGGCGGAGGAGAAAAAGGAGGCGYGLGSGAGRG-----GGAGAGAA 284
Db 873 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 932

OY 285 AAGAGGCGYCGGAGCGGCGYGLGSGAGRGGLGCGGAGAAAAAGAGCGGGLGSGGA 344
Db 933 SGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGS 992
OY 345 GAAAAAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 398
Db 993 GAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 1052
OY 399 ---GCGAGAAAAAGAGGAGGCGTGLGCGGCGYGLGCGGAGAAAAAGAGCGGCGVSSGA 455
Db 1053 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 1112
OY 456 SAASAAASRLSPASSRVSSAVSNLVAASPTNSALSTISINVSQIGASNGLSGCDV 515
Db 1113 GAGSAG-----AGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGMDPG----- 1159
OY 516 LIOALLGHHH 526
Db 1160 RYOLSGRHYH 1170

RESULT 10

US-08-477-509B-64
; Sequence 64, Application US/08477509B
; Patent No. 5770697
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W
; APPLICANT: Dorman, Mary A
; TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
; TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
; NUMBER OF SEQUENCES: 112
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albilton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,509B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-509B-64

Query Match 36.0%, Score 972, DB 1: Length 1177;
Best Local Similarity 44.1%, Pred. No. 3e-64;
Matches 243; Conservative 30; Mismatches 224; Indels 54; Gaps 17;

QY 17 GYGGLGGGAGGAGAAAAAAG-----GAGGGGCGGGLGSGAGRG--GCGAGAAAAAG 70
DB 633 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 692
QY 71 AGGCGGCGGAGS--GAGAGCGGAGGAGAAAAAGGCGGAGCGAGAGAAAAAGAGAG 129
DB 693 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 752
QY 130 GCGGGLGS---GAGAGSGSGGCGAGAAAAAG--GAGGCGGCGGAGS--GAGAGCGGAG 184
DB 753 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 812
QY 185 AAAAAAG---GAGGCGGCGGAG--GCGAGCGGCG--GLGS--GAGAGCGGAGGAGAAAAAG-- 237
DB 813 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 872
QY 238 ---GAGGCGGCGGAGAAAAAGAGGCGGCGGAGAGRG-----GCGAGAAAA 284
DB 873 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 932
QY 285 AAGGAGGCGGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 344
DB 933 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS 992
QY 345 GAAAAAGGAGGCGGAGGAG 398
DB 993 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 1052
QY 399 ---GCGAGAAAAAG 455
DB 1053 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 1112
QY 456 SAASAASRLSSPQASSRYSSAVSNLVAAGPTNSAALSTISNVVSIQAGSPGLSGCDV 515
DB 1113 GAGSGAG-----AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 1159
QY 516 LIQALLGHHH 526
DB 1160 RYQLSAGRHY 1170

RESULT 11
US-08-707-237A-35
Sequence 35, Application US/08707237A
Patent No. 5830713
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Capello, Joseph
APPLICANT: Criseman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
NUMBER OF SEQUENCES: 108
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/707,237A
FILING DATE: 03-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-10/WHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-707-237A-35

Query Match 36.0%, Score 972, DB 2: Length 1177;
Best Local Similarity 44.1%, Pred. No. 3e-64;
Matches 243; Conservative 30; Mismatches 224; Indels 54; Gaps 17;

QY 17 GYGGLGGGAGGAGAAAAAAG-----GAGGCGGCGGGLGSGAGRG--GCGAGAAAAAG 70
DB 633 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 692
QY 71 AGGCGGCGGAGS--GAGAGCGGAGGAGAAAAAGGCGGAGCGAGAGAAAAAGAGAG 129
DB 693 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 752
QY 130 GCGGGLGS---GAGAGSGSGGCGAGAAAAAG--GAGGCGGCGGAGS--GAGAGCGGAG 184
DB 753 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 812
QY 185 AAAAAAG---GAGGCGGCGGAG--GCGAGCGGCG--GLGS--GAGAGCGGAGGAGAAAAAG-- 237
DB 813 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 872
QY 238 ---GAGGCGGCGGAGAAAAAGAGGCGGCGGAGAGRG-----GCGAGAAAA 284
DB 873 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 932
QY 285 AAGGAGGCGGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 344
DB 933 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS 992
QY 345 GAAAAAGGAGGCGGAGGAG 398
DB 993 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 1052

OY 399 ---GOGAGAAAAGAAGCGGTGGTGGCCTGCAGCAAAAVAGACGCGTCGYVSGA 455
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1053 AGAGSAGACAGCACGACGACGACGAATGCAGACGACGACGACGACGACGACGASGA 1112

OY 456 SASASAARSRLSSPQASSRVSAVNLYASCPTNSALSTISNVYSOIGASNPLSCDV 515
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1113 GAGSGAG-----AGSAGAGSAGAGSAGNAAYGAGAGSAGAGSGAGAMDPC----- 1159

OY 516 LIOALLGHNNH 526
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1160 RYOALSAGRHY 1170

RESULT 12
US-08-482-085B-64
Sequence 64, Application US/08482085B
Patent No. 6018030

GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Richardson, Charles
APPLICANT: Chambers, James
APPLICANT: Causey, Stuart
APPLICANT: Pollock, Thomas J.
APPLICANT: Cappello, Joseph
APPLICANT: Cripsman, John W.

TITLE OF INVENTION: NO. 6018030el Peptides Comprising Repetitive
NUMBER OF SEQUENCES: Units of Amino Acids and DNA Sequences Encoding the Same
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fleher, Honbach, Teat, Albritton & Herbert
STREET: Four Embardadero Center, Suite 3400
City: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,085B
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993

ATTORNEY/AGENT INFORMATION:
NAME: Treccarlin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-482-085B-64

```

Query Match Similarity      36.0%: Score 972: DB 3: Length 1177:
Match Local Similarity     44.1%: Pred. No. 3e-64:
Matches 243: Conservative 30: Mismatches 224: Indels 54: Gaps 17:

QY 17 GYGGTGGGAGGAGGAGAAAAAAG-----GAGGGYGGTGGAGAGRG-CGGAGAAAAAAG 70
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 633 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAG 692
QY 71 AGCGTGTGCTG--CGAGRCGTGCGGAGAAAAACAGVCGCGTGGCGAGCGGAGAAAAAGCAQ 129
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 693 AGAGGCAAGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAAGSGAGSDA 752
QY 130 GCGYGLGSG--CGAGRCGCGCGGAGAAAAAAG-GAGCGYGGTGG--CGAGRCGTGGCGAG 184
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 753 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAG 812
QY 185 AAAAAAAG--GAGCGYGGTGG--CGAGCGGCGYGCAGAG--CGAGRCGTGGCGAGAAAAAG- 237
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAGAGSGAGA 872
QY 238 ----GAGCGTGGCGAGAAAAAAGCAGCGCGTGTGCTGCGAGRG-----CGAGAAAA 284
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 873 GSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAGAGSGAGAG 932
QY 285 AAGGAGCGCGYCGCGAGCGCGYGGTGGTGGAGRCGTGGCGGAGAAAAAAGCGCGTGGCGGA 344
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 933 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAGAGS 992
QY 345 GAAAAAAGCAGCGTGGTGGCGAGCGAGAA-----AAAAAAGGVCGCGYGGTGG--CGAGRG 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 993 GAGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 1052
QY 399 ----CGGAGAAAAAAGCAGAGCGGTGTGCTGGCGYGCAGTGGCGAGAAAAVGAAGCGYCGGSA 455
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1053 AGAGSGAGAGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 1112
QY 456 SAASAAASRLTSPQASSRVSAYSNLVASGPTNSAALSTISNVVSQLGASNPGLSGCDV 515
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1113 GAGSGAG-----AGSGAGAGSGAGAGSGAAGYGCAGAGSGAGAGSGAGAGSGAGAMPG----- 1159
QY 516 LIALLLGHHH 526
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1160 RYLSAGRYH 1170

RESULT 13
US-08-475-411A-31
: Sequence 311, Application US/08475411A
: Patent No. 6140072
: GENERAL INFORMATION:
: APPLICANT: Ferrari, Franco A.
: APPLICANT: Cappello, Joseph
: TITLE OF INVENTION: Functional Recombinantly Prepared
: TITLE OF INVENTION: Synthetic Protein Polymer
: NUMBER OF SEQUENCES: 119
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: US
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,411A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:

```



```

Db      813 AGSGAGAGSSGAGAGSGAGAGSGAGAGSGAAGCYGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 872
Oy      238 ----GAGCGGGLGCGAGACAAAAAAGAGAGCGGCGTGGGSGAGRG-----GCGAGAAA 284
Db      873 GSGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAAGAGSGAGAGSGAGAGSGAGAGSGAG 932
Oy      285 AAGGAGGCGGCGGAGCGGCGTGGGCGAGCGGGLGCGGAGAAAAAAGAGAGCGGCGGLGCGGA 344
Db      933 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAAGAGSGAGAGSGAGAGSGAGAGSGAG 992
Oy      345 GAAAAAAGAGAGCGGGLGCGAGCGAGAAA-----AAAAAAGAGCGGCGGCGGGLGGS-OGARG 398
Db      993 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 1052
Oy      399 ---GCGAGAAAAAAGAGAGCGGCGTGGGCGGCGAGGGLGCGGAGAAAAAAGAGAGCGGCGGSGA 455
Db      1053 AGAGSGAGAGSGAGAGSGAGAGSGAAGYGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 1112
Oy      456 SAAGAAASRLSSPGASRRSVSSVSNLVASGPTMSAALSTISINWVSQIGASRPLGSLGCDV 515
Db      1113 GAGSGAG-----AGSGAGAGSGAGAGSGAGAGSGAAGYGAAGAGSGAGAGSGAGAGSGAG 1158
Oy      516 LIQALLGHHH 526
Db      1160 RYGLAGAGRYHH 1170

RESULT 15
US-08-175-155-48
: Sequence 48, Application US/08175155
: Patent No. 5641648
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
APPLICANT: Cirlsman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: Methods for Preparing Synthetic
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fleht, Hohnach, Test, Albrition & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175.155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertam I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1059 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-175-155-48

```

Best Local Similarity 45.4%; Pred. NO. 1.4e-63;
Matches 244; Conservative 33; Mismatches 225; Indels 35; Gaps 18

QY 17 GYGGLGGGAGGCGAAGAAAAAG -GAGGGGTGGLGSQGAERG -GGCAAGAAAAAGAGGG 74
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 524 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS -GAGAAGAGAGSGAAGSGAGAG 582
QY 75 GYGGLGS -QGAGRGGTGGGCGAAGAAAAAG --GYDGGTLGGGCGAGGCGAAGAAAAAGAGGG 130
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
.Db 583 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 642
QY 131 GYGGLGS --QGAGRGGSGGCGAAGAAAAAGAGAGGGGTGGLGS -QGAGRGLGGGCGAAGAA 186
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 643 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 702
QY 187 AAAAAAG ---GAGGGGTGGLG ---GGGAGGGGTGGLGS -QGAGRGLGGGCGAAGAAAAAG -G 238
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 703 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 762
QY 239 AGGGGLGGCGAAGAAAAAGAGAGGGGTGGLGS -QGAGR -GGCAAGAAAAAGAGAGGGGTGG 296
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 763 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 822
QY 297 OGAGGGGTGGLGS -QGAGRGLGGGCGAAGAAAAAGAGAGGGGTGGLGGCGAAGAAAAAGAG 355
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 823 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 882
QY 356 QGGTGGGAGAGAGAAAAAGAAAAAGGVRGGGTGGLGSQGAERG -GGCAGAAAAAGAGAG 414
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 883 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS -GAGAAGSGAGAGSGAGAGSGAGA 941
QY 415 G-----GTGGLGGGCVAGAGLGGCGAAGAAAAAGAGGGGTGVTGSGASAASAKSLSPQ 469
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 942 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA 1001
QY 470 ASRSVSSAVSVLSASPSTNSALSTPTSNVVVGIGASNPGAGSCGVLLDALLGHHH 526
Db 1002 GSAGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGMDP-----RVLDSGRHY 1052

Search completed: July 3, 2001, 14:56:45
Job time: 535 sec

Query match 35.68; Score 962.5; DB 1; Length 1059;

Db	438	AAAAAAGGAGGGGCGYGGGLGAGGAGRGGGGCAAAAAAGGAGGCGYGGGLGSGAGRGGGGAGGAAA	497
Qy	337	-----GGLGCGGAGAAAAAAGGAGGGGLGGCGAG	365
Db	498	AAAVAGGCGEGIRGCGAGGCGYGGGLSGGSGRGLGGCGGAGAAAAAAGGAGGGGLGGCGAG	557
Qy	366	CGAGAAAAAAGGAGGVRGCGYGGGLGSGCAGRGGGGAGAAAAAAGGAGGCGGTGGLGGCGY	425
Db	558	CGAG-----AAAAAAGGAGVRGGGTGGGLGSGCAGRGCGCAGAAAAAAGGAGGGGTGGLGGCGY	613
Qy	426	GAGGLGCGGAGAAAAVGAAGGCGYGGVSGCASAASAAASRLSPQASRRVSSAVSNLVASG	485
Db	614	GRGGLGCGGAGAAAAAGGAGGCGYGGVSGCASAASAAASRLSPQASRLSSAVSNLVATG	673
Qy	486	PTNSAALSTTISNVVSQIGAS	506
Db	674	PTNSAALSTTISNVVSQIGAS	694

RESULT 2
A44112
spidroin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)
N:Alternate names: silk fibroin, dragline
C:Species: Nephila clavipes
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Mar-1998
C:Accession: A44112, S27824
R:Hilman, M.B.; Lewis, R.V.
J. Biol. Chem. 267, 19320-19324, 1992
A:Title: Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes
A:Reference number: A44112; MUID:92406876
A:Accession: A44112
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-627 <HI>
A:Cross-references: GB:M92913; NID:g159713; PTD:g159714
A:Note: sequence extracted from NCBI backbone (NCBI:113893)
R:Hilman, M.B.; Lewis, R.V.
submitted to the EMBL Data Library, May 1992
A:Description: Isolation of a clone encoding a second dragline silk fibroin: Nephila clavipes
A:Reference number: S27824
A:Accession: S27824
A:Molecule type: mRNA
A:Residues: 19-627 <HI>
A:Cross-references: EMBL:M92913

[illegible]

Oy	316	GLGGGCGA	AAAAAAAAAGAGCGGCGGGGAG-----	AAAAAAAAG-----	GAGGCGG	358	
Db	343	GGGGTGP	SASASAAAAAPGQGGCGGCGPGCGQGP	SGPSSASAAAAAAAAGPGGTGPGQGS	402		
Oy	359	LGCGGAGG-----	GAGAAAAAAAAGGVRGGCGYGLGSGAG-----	RGCGGAG-AAAA	407		
Db	403	PGGYAP	CGGQGP	SGPSSASAAAAAAG--PGGYGP--CGQGP	PGYAPCGQGP	459	
Oy	408	AAGGAGGCGTC-----	GLGGGCV-----	CAGGLG-----	GGGAGAAAAGGAGCGG	447	
Db	460	AAAAAGP	GGYPPA	QGGSPETIAASASAGGGGCGP	AQCGPAGTGP	SAVAASAGAGSAG	519
Oy	448	YCGVSG	SASASASRLSP	QASRSVSAVSNLVAASPTNSALST	ISIVVSOIGASN	507	
Db	520	YG----	PGSQASASASRLASPD	GARVAASVNSLVSSGPTSSAALS	SVISNAVSOIGASN	575	
Oy	508	PGLSGC	DVLIAL	521			
Db	576	PGLSGC	DVLIAL	589			

```

RESULT 3
T31328
fibroin - Chinese oak silkworm
C:Species: Antherea pernyi (Chinese oak silkworm)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999
C:Accession: T31328
R:Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, A
A:Reference number: Z20995
A:Accession: T31328
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2639 <SEZ>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AC32606.1
A:Genetics:
A:introns: 14/3

```

Query Match	Similarity	35.9%	Score 968.5	DB 2:	Length 2639
Best Local	Similarity	34.7%	Pred. No. 1.5e-45		
Matches	287	Conservative	30	Mismatches	160
				Indels	351
				Gaps	26

Oy	231	GAAGAAGAGCGGCGGCG-----	-----	QCAGAAAAAAGCAGCGGCTGCGGCGAG	273
Db	1200	AAAAAAGSAGAGAGCGTGWGDCGTGSDSAAAAAAAAAAAAAAAAA	-----	SGAGSGGCGTGTGCGTGS	1259
Oy	274	RCGCGAGCAAAAAAGCAGCGCGT- GCGCGAGCGCGT	-----	GLCSGCGAGG- -	315
Db	1260	SAAAAAAAAAAAGSAGAGCGTGTGWDGCTGSDSAAAAAAAAAAAA	-----	ASGAGGCGGCGT	1319
Oy	316	-----	-----	GLGCGGAG- -	323
Db	1320	SGSGSAAAAAAAAAARRAGHDBRAGSAAAAAAAAAAAAAAAAA	-----	ASGAGSGGCGTGWGDCGTG	1379
Oy	324	-----	-----	AGAAAAAAGCAGCGGL- GCGAG	365
Db	1380	SDSAAAAAAAAAAGSAGSGGTGCTGSDSAAAAAAAAAAAAAAAAA	-----	AGAGGCGGTGCGT	1439
Oy	366	CG-----	-----	AGAAAAAAGAGTGTGCGTGLGSGGAGCGGCGAGAAAGAGCGGCT	417
Db	1440	DGCTGSDSAAAAAAAAAAGSAGSGGTGCTGSDSAAAAAAAAAAAA	-----	ASGAGGCGT	1499
Oy	418	GGLGCG-----	-----	GVGAGGLG-----	447
Db	1500	GCTGSDSAAAAAAAAAAGSAGGAGCGTGTGWDGCTGCTGSA	-----	AAAAAAGAGGAGG	1559
Oy	448	T--GCTGCGAGSAAASRLSPQASRTVSASVNLVASGPTNPA	-----	ALS	493
Db	1560	RGDGTGCTGSSAAAAAAAAAARRAGHDBRAGSAAAAAAAAAAAA	-----	AS	1607

RESULT 4
 F70806 hypochelical glycine-rich protein RV3508 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: F70806
 R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Dellin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtz, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Accession number: A70500; MUID:98295987
 A:Reference: F70806
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1901 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17745.1; PID:g2924444
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3508
 C:Superfamily: collagen alpha 1(IV) chain

QY	198	GYGLIGGCG-----AGGGYGTGGLSGG-----AGGGGLIGCG--GAGGAAA	235
Db	902	GAGGAGGAGGADNNFNFGGCGGAGCGGGGGLGCASTTSTIMANGAGGNGGTGGKGGAGGAGT	961
QY	236	AGGAGGCGGLGCGCGA-----GAAAAAAGAGAGCGGTYGGLSGGCA-----G	273
Db	962	LGVGSSGCTGGDDGDAGSGGGCGTCCGAAGAAGGGGNGRGGCGDGAISGLGLGSPDGG	1021
QY	274	RGCGAGAAAAAAGGA--GGGTGGCGAGCGGTGGLSGGAG--RGLIGCG--GAGAAA----	326
Db	1022	GGCGGAGGAGGAGGAGGAGGAGG--GDDGDDATGAAGTGLMDGVTGGDGGAGGAAANGC	1080
QY	327	-----AAAAAGG-----GGGGLSGGCG-----AGAAAA--	350
Db	1081	NAGVLTAKKADGGAAGNGGAGGAGGAGGAGMDNPNFGCGGAGCGCGGGLGGAATTSTIN	1140
QY	351	-----AGGAGGGGGLSGG-----ACGGAGAAAAAAGAAAGVROGG	386
Db	1141	ANGGAGGNGCTGCKGKGAGGAGTTLVGCGSGGTGGDGDAGSGGGGCGTGGACAGCGGNGG	1200
QY	387	YGGGLSGGCA-----GRRGGCGAGAAAAAGG--AGGGGTGGLGGGCV-----	425
Db	1201	VGGDGGBEASGLGLGLSGFPDGGGGGGGAGGSGAGAGGINGAGGAGGTTGGAGGDDGAPATLI	1260
QY	426	-----GAGGLGGGCGAGAAAAVGA-----GGGTYGCVSSGASAASA	460
Db	1261	GGPDGGDGGCGGTGGDGGNAGFGAGVPGDGGDGGNAGFGAGVPGDGTGGTGAAGAGGA	1320
QY	461	AASRLSSQASRRVSSAVSNLVAAGPNSAALSTTISNYSQIGASNPRLSGCD	514
Db	1321	GADGPGSTIDGGGAGGAGGCGGCGKGTINTSTGLSAASBGDGGNGCAGAGAGNGCD	1374

hypothetical glycine-rich protein RV3512 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70807
R:Colte, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
Comor, R.; Davies, R.; Dellin, K.; Fellwieser, S.; Hamlin, N.; Holroyd, R.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, K.; Stulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A705007; MUID:98295987
A:Accession: B70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 11079 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17749.1; PID:g292
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3512
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

OY	154	AAAAAGACGGCGTGGTGGGCGAGRGGLGCG-----AGAAAAAAAGC-ACGGCT	199
Db	490	GGAAGDGGCGCGGAGGAGGAGCGGAGGAGGTGCNGNITGTAGTAAGAANGAAGGGA	549
OY	200	GGTGGGAGGCGGYGGLSSOG-----AGRGLGSGO---A	230
Db	550	GGCGGTGGGTGGGCGAGGDDGAGGTGGDRTYGGCTVPAAGSOGCGGNAGGGGAGGGGADG	609
OY	231	GAATAAAGACGGCGTG-----GQAGAAAAAAGACGCGGTGGLGSGAGRGCG-GAGA	281
Db	610	GSGGDGDGAGTGGNGMGNGMNRNSGNTGGAGNGCGGAGGAGGAGGAGGCGGTGGNGGAGG	669
OY	282	AAAAG-----GAGCGGTGGCG--ACGGCTG--GLSSQCAGRGCLGCGAGAAAAAAA	331
Db	670	DACGAGMGNGNCGTGNGNGNGGCIAGMGNGAGTGSNGNGNSGSGNGMAOMGNSG	729
OY	332	GGADGGGTCGGGAGAAAAAAGC-----ACGGTGGGAGACGAG-A-AAAAAAAAG	380
Db	730	TGSDGDGAGGNGCAATGTGTGGDGLTGTGTTGSSGCTGGDGCGNGABDNMTAQAG	789
OY	381	GVRCGYGXYGLSSQCAGRGCGAGAAAAAAGACGCGTGTGLGGCY-----	425
Db	790	G--DGGNGCGDGFEGGAGAGGCGGLTAGANCTGGGAGAGCGDGCNGCALIGHBPLTDPCGNG	847
OY	426	GAGTGGGAGAAAAVGA-----GCGGTGGV-----SCASAASAAA	462
Db	848	GTGNGCTGTGTGGAGTGGTGGTGGDGGNGNGGTGGBGCEVGAGGCTGAA	899

RESULT 6
D70807
hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fieldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1489 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g2924454
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3514
A:Superfamily: collagen alpha 1(IV) chain

Query Match	33.8%	Score 912.5;	DB 2;	Length 1489;
Best Local Similarity	39.1%	Pred. No. 1e-42;		
Matches 263;	Conservative 19;	Mismatches 202;	Indels 189;	Gaps 34;
Oy	2	ASMTGGQOMGRIRINGYGLGGGQAGQ--GAGAAAAAAGAGAGCGYGLGSGAGRGGO-	59	
Db	178	AGCAGGMLFG---VEGAGAGVGAGGGGTGGAGGPGLLINGGGAGVGAGGAGGTGAGGGA	234	
Oy	60	---GAGAAAAA--GG-----AGCGYGLGSGQ-----AGRGGLGGQ--GAGAAAAAGGY	103	
Db	235	ELLFFAGAGGAGGAGTGGCGATGCTGTGHHGAGGDDGGLWLPAGGAGGAGGAGGASDSCGA	294	
Oy	104	--GCGGLGGGQAGGAGQ--AAAAAAGAGAGCGYGLGSGQ-----AGRGSGGGO-	148	
Db	295	LGGTGTGGTGGAGAGGAGRGALLLGAAGCGGGLGAGAGCGGCTGAGAGGDDVLGGVGGTGGKG	354	
Oy	149	-----GAGAAA-----AAAGAGAGCGGYGLGSGAGRGGLGSGGAGAAAAAAGAGG	195	
Db	355	GVGGVAGVAGGAGGAGGAGGOLFSSAGAGAGNAGVGGAGGQG--GDGAGGAGGAGDADDPGATGCTGTG	413	

OY	196	-QGGYGLGSGQ-----AGCGYGLSGQA-----GRGLGSGQ-----AG	231
Db	414	FAGGAGGAGGAGGSSGAGCTGNSGAGCGGAGGAGAGADNPRTIGTGGDGTGGAGG	473
OY	232	AAAAAGGAGGCGGLGQ-GAGAAAAAAGGAGGCGYGLGSGQA-----GRG	275
Db	474	AGGAGGAAGTGTGTGMITGTGTMAGCYGAGCGGGDGGAGAGADNPATGATGCGFAGGAG	533
OY	276	GQ-GAGAAAAAAGGAGGCGYGGCGAGCGYGLGSGQA-----GRGLGSGQ--ACAAAA	328
Db	534	GAGGAGGSSGAGCTGNSGAGGAGCTG-CGGGAGGAGGAGADNPRTIGTGGDGTGTAAGAG	592
OY	329	AAAGGAGCGGGLGQ-GAGAAAAAAGGAGCGGLGSGQ-----ACGAGA	370
Db	593	GAGGAAGTGGGAGMITGTGTMAGCYGAGGCGGCGDGGAGAGADNPATGTGTGAGAGAGA	652
OY	371	AAAAAAAAGCV-----RGGYGLGSGQA-----GRGCGAGAAAAAG	410
Db	653	GKAGSSSAGCTNSGAGCTGRTGGAGAGADNPRTIGTGGDGTGCAAGGAGAG	712
OY	411	G-AGGCGTGL-----GGCGV-----GAGGLGQ-GAGAAAAAGA-----	443
Db	713	GAAGTGGTGMITGTGTMAGCYGAGGAGSSGAGCTGNSGAGCGTDDCGAGGAGAGADNP	772
OY	444	--GGGYGVSAGSAASAAASRLSSPOASRVSSAYSNLVSAGPTNSAALSTISNVYS	501
Db	773	IGTGGDGTGGAAGAGAGAG-----AAGTGTGMITTTGNA-----	810
OY	502	QIGASNPGLSGD	514
Db	811	--GVGAGGCGGD	821

RESULT 7
E70806
hypothetical glycine-rich protein RV3507 - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70806
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: E70806
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1381 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:93261554; PIDN:CAA17744.1; PID:9292
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3507
C:Superfamily: collagen alpha 1(IV) chain

Query Match	33.0%	Score 892	DB 2:	Length 1381
Best Local Similarity	34.7%	Pred. No. 1.3e+1		
Matches 258	Conservative 25	Mismatches 218	Indels 242	Gaps 32

QY	2	ASWTGQGMGRIRIRISYGG	GLGOGAGOGAGAAAAAGAGGCGY	-----	GGLSOGA	54
		:	:	:	:	
DB	149	AGOSGN-----	-GGSAGLMNGAGGAGGSGGAAGNGNGGCLFPAAGTGTGTGGA			201
QY	55	-----	GRGOGAGAAAAAGAGOGG	YGG	GLGSQAGRGGLGOGAGAA	-AAAGV -GQG 106
DB	202	PGAMGTGTGNG -GNALLIGGG	GLGAGAGMGGTGGTGGTGGNGG	GNALLIGAGG	VGAG	260
QY	107	GLGOGAGOGAGAAAAAGC	-----	AGGGYGLGSQA	-----	GRGSSG 147
		:	:	:	:	
DB	261	GIGGGTGTGAGGAAGGTGTG	NGGAGGLFNNGGDGGAGGGG	GTGGGAADAAASAGTGTGGG		320

[illegible]

RESULT 10
S01820
glycine-rich cell wall protein 1.8 precursor - kidney bean
C:Species: *Phaseolus vulgaris* (Kidney bean)
C:Date: 30-Sep-1999 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C:Accession: S01820
R:Keller, B.; Sauer, N.; Lamb, C.J.
EMBO J. 7, 3625-3633, 1988
A:Title: Glycine-rich cell wall proteins in bean: gene structure and association of the
A:Reference number: S01820; MUID:89091109
A:Accession: S01820
A:Molecule type: DNA
A:Residues: 1-465 <KEL>
A:Cross-references: EMBL:X13596; NID:g21002; PIDN:CAA1932.1; PID:g21003
C:Comment: This protein is enriched in the cell wall fraction of young hypocotyls and ov-
tlys.
C:Comment: Much of the sequence consists of tandemly repeated 22-residue segments with 1
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8
C:Keywords: cell wall; structural protein; tandem repeat
C:i-30/Domain: signal sequence #status predicted <SIG>
C:i31-465/Product: glycine-rich cell wall protein 1.8 #status predicted <MAT>

Query Match	32.1%	Score 866.5;	DB 1.	Length 465;
Best Local Similarity	47.8%	Pred. No. 1.2e-40;		
Matches 218; Conservative	8;	Mismatches 177;	Indels 53;	Gaps 19;

[illegible][illegible]

RESULT 11
A70896
hypothetical glycine-rich protein Rv1091 - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: A70896
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
R.: Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajadream, W.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: A70896
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <COL>
A:Cross-references: GB:AL021897; GB:AL123456; NID:93256022; PIDN:CAAI7207.1; PID:e125
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1091
C:Superfamily: unassigned collagens

Query Match	31.8%	Score 858;	DB 2;	Length 853;
Best Local Similarity	42.8%	Pred. No. 5.9e-40;		
Matches 232; Conservative	18;	Mismatches 184;	Indels 108;	Gaps 26;

QY 7 GCGMGRIRIRGTCGTCGCGACGACAGAAAAAAGAGAGCGCTGCTGCGA----GRGGGAGA 63
 Db 292 GGGGGHALLMTGAGGAGCGNCGSGCTGAGAGGTGACGCGNGAGCGGGCTGCTGTCGAGCGG 351
 QY 64 AAA-----AAGGAGCGCTGCTGCTGCGAGCGTCGCG-----GAGAAAAAGTGGCGCT 108
 Db 352 AAAGMGTAAGNGVSSSSGGGAGCTGAGCGDGC--GAGGAGGNAITMTGTGAGAGAG--GDGGA 408
 QY 109 GCGAGCGAGAAAAAAGCA----GCGCTGCTGCG-----AGRGSGCGGA----- 150
 Db 409 GGAGGCGGSGCTGSMNNGAGGAGGDSRGCTGAGAGGCGAAILTGTGCHGCGDGAGCAAAVKG 468
 QY 151 --GAAAAAAGGAGCGCTGCTGCTGCGAGCGCTGCGCGAGAAA-----AAAA 191
 Db 469 GDDGAAAGTGTGAGCGRRGAGC---GSGGSGDDGGGAAAGPAGMTLTGDDGAGCGNGGAAAA 524
 QY 192 GGGAG--GGCTGCTGCGGAGCGGCTGCTGCTGCGA-----GRGCTGCGCGAGAAAAAGCA---- 239
 Db 525 GGAAGCGAGGGGGGCGGCG--GNGGNGGNGGNGATGTGCTGTGCGAGAGCGCATATGAGAGAGAGNV 583
 QY 240 -----GCGCTGCG--GAGAAAAAAGAGCGCTGCTG--GCGAG--RGCGAGAGAAAAAGG 288
 Db 584 STNGGCTGTGNGCTGTGCTGCGGAGAGGNNGLLTGTGAGCGHGAAGCGAGTGTFTSSDGG 643
 QY 289 AGCGGTGTGCGAGCGCTGCTGCTGCGAGCTGCTGCG--GAGAAAAAAGAGAGCGCTGCGCGAGAA 347

Db 644 A--GGDGDG--GNGAGGTGGLTFGAGGNGGPGGSGGAADIGNGGAGNGG--GTDTGNGGN 699

Oy 348 AAAAGAGGGGLGGGAGGGA-----GAAAAAAAAGCVRGGGTG--GLGSQG-- 394

Db 700 GGSGGAGSGGGDGGAGGNGAMLPGNGGAGGGGGKCGAGGGGLGGGSFGLPLGLNGSGGD 759

Oy 395 AGRGGGGA--GAAAAAAGGAGCGGTGGLAGCGV--GAGGLGGGCGAGAAAAAGAGCGGTGG 450

Db 760 GGDGCGNGAPGCVLVYKNGGAGGCGSSGGTGGPGPANGAGAGKGGDGDALIDGGGNGNGG 819

Oy 451 VG 452

Db 820 AG 821

RESULT 12

Hypothetical glycine-rich protein RV0834c - *Mycobacterium tuberculosis* (strain H37Rv)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence,reviston 17-Jul-1998 #text,change 20-Jun-2000
C:Accession: B70812
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentsel, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, S.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: B70812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1,882 <COL>
A:Cross-references: GB:AL022004; GB:AL123456; MID:g3261550; PIDN:CAAI7640.1; PID:g2916888
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV0834c
A:Superfamily: elastin

Query Match	31.88;	Score 857.5;	DB 2;	Length 882;
Best Local Similarly	35.28;	Pred. No. 6.5e-40;		
Matches 247; Conservative	20;	Mismatches 226;	Indels 209;	Gaps 29

OY	5	TGGGQMGKIRTRRGVGGGLGGGAGCGAGAAAAAAGA----	GQGCVGLGSQGAERGGGA	61
Db	157	TGGAG-GAAAGLLGGCTGGAG---GTGASGGKCGTGTWLWSSGCGAGSGSGSGGGG		212
OY	62	-----GAAAAAGCAGCGCGVLGSOG-----	AGRGLGC---QCGAAAAAGG	102
Db	213	NALMFGLGNCGACGAMASGVNGVGAGAGALVAIGAGAGGAATTTTGAGCGAGS		272
OY	103	-----VGQGLGGGOGA----GCGAGAAAAAAGA-----	-GGGYGGLGSQGA	141
Db	273	NALGLFLGLGSGGQGDPSAMGSGGAGGSGGASPFIDIGIGAGGHGGAETN-G		331
OY	142	RGSQSGG-----AGAAAAAGAGAG-----	-QSGY	165
Db	332	AGGAGGAGGSSGTYFALDLSWGAGGNGCAATTGTGGAGGTGGFAVPADFTGFGANYGGA		391
OY	166	GGLSQGAGRSGGLCGGAGAAAAAACGACGGCGYGCLGCGAGOGGYGGLGSQAG--RGG		224
Db	392	GGLGGAATGAGCTGCTGGYVAGAGPAAICVYGAGAGAGGAATTETGIGTAGAGLGVGLLG		451
OY	225	LGGOGA-GAAAAAGCAGGAGGGLGCGAGAAAAAAGAG-----	-QGGYGGILGSOG----	271
Db	452	AGGAGGPCGMASSAGSGGAGCTGCDALGLIGAGIGGVGVGGAAMDYTGNGANGAGSGGTLL		511
OY	272	--AGRGGGAGAAAAAAAAGAAGOGYGGGAGAGOGGYGGLGSQAG---	RGGLGQGGAGA	324
Db	512	GGYVGAGGHGGGASVCTGCSGAGAGSDGRFPVAGAGNGMNACTGYVNVANANGNGNSAATGA		571
OY	325	AAA---AAAAGG---AGCGGLGGO-----GAGAA-----		347

Db	572	LAAVGAGAACGADATSGTGTGGCGAGCSARCLIFALGAGAGAAGDASTVGVGPGGTGTT	6311
Oy	348	-----AAAAGGAG-GGGTGGCA-----	
Db	632	AASPGLIALTCGAACAGGAGTSRTATGAGAGDDYCFEGIVLVGLTFGCAGACGAAITDGA	692d
Oy	384	OGGTGGTCTGCGAG-----RGCGCAGAAAAAAGAGCGGTGTGCG-----OCVGA	427
Db	692	TGGAGGFFCGACAGIANELGFSEVLHGGAG-GAGGTATRTGGKMGAGGGGGGLSSPVTLLIGI	750c
Oy	488	GGLGGCGAGAAAVAGCGGCGYGVSCHSAASAASRLSPQASRVSSAVSLVNASGPT	487f
Db	751	GGAGGGDGAGLGVLTG-GMCGDPGDGGEAVANGIAV-----	
Oy	488	NSAALSTITSINVSQIG-----ASNPLDSGCVLI	517
Db	797	GNCGAGNGGAGLGLVGVGNGNAGCTFCGANTGTGNGGNDTTI	838

RESULT 13

hypothetical glycine-rich protein RV0833 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70812
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
J.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: A70812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-749 <CDL>
A:Cross-references: GB:AL022004; GB:AL123456; NID:93261550; PIDN:CAAI639.1; PID:g2931
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0833
C:Superfamily: elastin

Query Match	31.5%	Score 850;	DB 2;	Length 745;
Best Local Similarity	36.5%	Pred. No. 1.4e-39;		
Matches 241; Conservative	21;	Mismatches 223;	Indels 176;	Gaps 25;

```

OY 7 GCGMKRIRIRGCGLGCG-----GAGCGAGNAAAAAAAAAGCA-GCGCGYGGLSGCGAGRG 57
Db 20 GGPALLILIVGAGGAGCGGSAVAGYIGGAGCGAGCAALLFLGAGAGGAGAGSGSGSGAAGGAGG 79
OY 58 GGGAGCAAAAAAGAGAGGCGGCGGLGSGGAG-----RGGLGGGCGAGAA-----98
Db 80 AGGAGCGFLPASGSGSGCGFGCFASITGTGGAGGTGAGAGGLPASGCGGTGGGAGSGCGTGVGT 139
OY 99 -AAGG-----GCGGLGCGGAGCGAGAAAAAAGAGAGGCGGCGGLGSGGAGRGSGCGGAGAA 153
Db 140 GGAAGGLFASGGAGGAGGSGCGGTGTGAGGTGAGAGGLFGAGAGGGLGGGQHHTGGHGGAGGSAG 159
OY 154 AAAAGGAGCGGCGGLGGLSGG -AGRGGLGCG-----GAGCAAAAA-----AGG 193
Db 200 LLALEDGAGCAGCGAMTTGTGTGAGAGAGGAKALLFGSGGAGGSGCGAMCTFBDTNGSGAGCG 259
OY 194 A-----GGGCGYGGLGCGGGA -GCGGCGGLGSGGAGRG--GLGCGG -AGAAAAAGGAG 240
Db 260 AGGKAGLLFGSGGAGAGSGSGAGGFRNGSTGTGGAGGAGGAGGLIGNGNGNGSGGTSVATGGAG 319
OY 241 OGGGLGCGGAGAA-----AAAGCAGCGGCGGL-----267
Db 320 NGGAGGAGGAGGAGLIGNGNGSGCGMGDAPGCTGVGCGILLLGIDGANNAPASTNP.LHTAQ 379
OY 268 -----GSQGA-----GRGCGGAGAAAAAAGG 288

```

D_b QQALAAVNAPELQAVTGRPLIGNGANGAPGSGAPGGHGWLEFGGGGTGGSSVSGAGGDGG 439
 380 QQAALAAVNAPELQAVTGRPLIGNGANGAPGSGAPGGHGWLEFGGGGTGGSSVSGAGGDGG 439
 OY A-----GGGGGGGAGGGGYGLGSOGA-----GRGGLGSGOGAAGAAAAAAGA- 334
 289 A-----GGGGGGGAGGGGYGLGSOGA-----GRGGLGSGOGAAGAAAAAAGA- 334
 Db AGGILFAGAGGAGAGCAVTCTGCATGSGSAGAGCAllLPAGAGGAGGSSGIGGFPAAAGAG 499
 440 AGGILFAGAGGAGAGCAVTCTGCATGSGSAGAGCAllLPAGAGGAGGSSGIGGFPAAAGAG 499
 OY GGGGLGGGAGNAAAAAAAAAGAG--QGGLGGGAGGAG--AAAAAAAAAAGVRRGGYGGL 390
 335 GGGGLGGGAGNAAAAAAAAAGAG--QGGLGGGAGGAG--AAAAAAAAAAGVRRGGYGGL 390
 Db GPGGAGGLEPNCGGAGAGGSGSVSGAGGEGGAGGCLPAGCGAGAGGSGNNVGAGGA 559
 500 GPGGAGGLEPNCGGAGAGGSGSVSGAGGEGGAGGCLPAGCGAGAGGSGNNVGAGGA 559
 OY GSQG-----AGRGGGAGAAAAAAGAG-----QGCTGTLGGGCVGAGGLGGGGA 435
 391 GSQG-----AGRGGGAGAAAAAAGAG-----QGCTGTLGGGCVGAGGLGGGGA 435
 Db GGCVGGLFPGAGGAGGSGSGGGSVAIBDSDAGNGLLPLTAGAGAGGGGGGFTGGAGGPGG 619
 560 GGCVGGLFPGAGGAGGSGSGGGSVAIBDSDAGNGLLPLTAGAGAGGGGGGFTGGAGGPGG 619
 OY GAATAVGA-GGGGTGGVC---SGASAAASAASRL-SSPQASRVSAAVSNLYASGPRTMSA 490
 436 GAATAVGA-GGGGTGGVC---SGASAAASAASRL-SSPQASRVSAAVSNLYASGPRTMSA 490
 Db DAGLLTVGSGGVGAGGfGLTTTGPCAAAGDGALLFGSGGAGAGGSGTRTDLGAGAGAGK 679
 620 DAGLLTVGSGGVGAGGfGLTTTGPCAAAGDGALLFGSGGAGAGGSGTRTDLGAGAGAGK 679
 OY A 491
 491 A 491
 Db 680 A 680

RESULT 14

E70917

hypothetical glycine-rich protein RV1450c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: E70917

R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: E70917

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1329 <COL>

A:Cross-references: GB:255844; GB:AL123456; NID:93250713; PIDN:CAB09271.1; PID:92131046

A:Experimental source: strain H37RV

C:Genetics:

C:Gene: RV1450c

C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match	31.4%	Score 847	DB 2	Length 1329
Best Local Similarity	38.9%	Pred. No. 3.3e-39		
Matches 243	Conservative 16	Mismatches 196	Indels 170	Gaps 27

QY	5	TCGQOMGRIRIRYGSGL--	GGGAGGCGAGAAAAAAGGAGCGGCGTGLGSOGAGRGGCGA	61
		: :	: :	: :
DB	409	SGGKAGNGNGAGGAGGLVNGGAGGAGCAGNAPGAPPSGGDPNGGGGGAGAGG--GKGGDS		466
QY	62	GAAAAAAGGAGCGGCGTGLGSOGA--GRTGLGGGAGAAAAAGGCGGGLGCGGAGGCGAG--		118
DB	467	GAGADGGAGGAGGAGGNGNGNGATGATGLNGAGADTGD--GKGGNGAGGCGGAGGGO		524
QY	119	-----AAAAAAGGAGGCGTGLGSOGA--GRTGGGCGGAGAAAAAAGGAGGCG		163
		: :	: :	: :
DB	525	GCKALATHTQDSSMGAGGAGCGAGCMGDDCGNGCKGTFTDNGGDSVGGNGNGSGSRGIG		584
QY	164	GTGGGLGSOGA-----GRTGLGGGAGAAAAAAGGA--GGGCGTGLGSOGA		207
		: :	: :	: :
DB	585	GAGGTGAGASTAGADGARGATPTSGGNGGTGNGAATVAGAGAGGKGGGNGGLVGN--		643
QY	208	GGGCGTGLGSOG-----AGRTGLGCG--GAGAAAAAG-----AGGCGGLGCGA		249
		: :	: :	: :
DB	644	GAGCGKGGDDMACVAGSSPTTAGESTSGGNGCAGGAGGAGGRCGDPGCGGTGAGGAGNAA		703
QY	250	GAAAAAAGGA--GGGCGTGLGSOGAGRGGGA-----GAAAA		284

Db	704	NGAATTTGAGCGGGCHGCPGAACG-GNCGGGGPGCGLANTLFGONGTGGVGSggkKGAGG	762
QY	285	AAG-----GAGGGGYGGG-----AGGGGYGGGGS---OGA-----GRG	315
Db	763	LAGGGGNCANGCFAGCDGNHGSGNGGNPCGAGGGGGSGGAGSTFTPAKKAHGFPTSSGDG	822
QY	316	GLGGGQ-----AGAAAAAAAAGCAGGGGGLGGGQA-----	344
Db	823	GDDGNGGNSQYVVGGNGGGDCGNGNGGNSAGCTGCNGCRGGDGAFFGMSANATNPENCPNGN	882
QY	345	---GAAAAAAAGCAG-QGGGLGGCAGCGAGAAAAAAAAGGVR-----QGTYGGLG	391
Db	883	PGGNGGAGAGGAGCAGCLINGNGGAGGNGGLGGTGCGAAGANCNVAAVGAPCGPCCAGCHGGAG	942
QY	392	SQG--AGRGGQ----GAGAATAAAGAGCGGTGGLGGGQGVG----AGTLGGCGACAAAA	440
Db	943	GNGGAGGNGGGQVYSDGAGGAGGAGGDDGAFBPDANGSNGOAGAFAGGGGGRRGGDGNA	1002
QY	441	VGAGGGGYYGVSASAAASAAASRL	465
Db	1003	GNAAGGPGGTGSTAGKAGPAGSILL	1027

RESULT 15
F70963
hypothetical glycine-rich protein RV2634C - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: F70963
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gellies, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R., Sulston, J.E., Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: F70963
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-778 <COL>
A:Cross-references: GB:280225; GB:AL123456; NID:g3242265; PIDN:CAM02341.1; PID:e26639
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2634C
C:Superfamily: unassigned collagens

[illegible]

D _b	505	GVCDAFGGDGDIGGDDGNALGAAGNGGTGTGAGGNGRCGLITLNGGAGGAGGTGGTGGG	564
O _y	281	AAAAAAGSAGGGCGVGGGG-----AGCGGVGCLTSOG-----AGRGLTGQGGAGAA	326
D _b	555	CAAGACGCY--GGAGGEBLTDGACTABEGCTGGTGLCGVGTTGGMGSGSGVGGAGAGS	622
O _y	337	AAAAAGCAGCGGCLGGQGAGAAAAAAGGAGGCGGLCGQGAGAGCAAAAAAAAAAGTVHQQG	386
D _b	623	LIGLGCGGAGGAGVCGTG-----GIGGIGGAGGNGAGAGCTTTGGATTTGGGGTGG	674
O _y	387	YGGLGSOG--AGRGQ--GAGAA-----AAAAGCAGCGGT---GGTGGGCVGAGCLGCG	433
D _b	675	VGGAGGTGTGTGAGAGTGTGSSGACGLTGMAGAAGCTAGCTTCGGCGGLCGQG--GNGGNGT	733
O _y	434	GA-----GAAAAVGAGCGGTGCVGSCASNAASAARLSPP	468
D _p	734	GATGGGCGGFALG--GNGAGCAGAGSPCGCSSSIOIQNMKP	771

Search completed: July 3, 2001, 14:59:45
Job time: 370 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:47 ; Search time 37.35 Seconds
(Without alignments)
484,254 Million cell updates/sec

Title: US-09-490-291-8
Perfect score: 2700

Sequence: 1 MASMTGGCGMGRIRIRIGYG.....GLSGCDVLIOALGHHHHH 528

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	84.2	747	1 SPDL_NEPCL	P19837 nephila cla
2	1148	42.5	627	1 SPDL_NEPCL	P46802 nephila cla
3	952.5	35.3	5263	1 FBOL_BOMMO	P05780 bombyx mori
4	935	34.6	1901	1 YZ08_MYCTU	O53533 mycobacteri
5	866.5	32.1	465	1 GRP2_PHAUV	P10456 phaeoculus v
6	844	31.3	778	1 YQ34_MYCTU	P71933 mycobacteri
7	839.5	31.1	914	1 WA22_MYCTU	O06794 mycobacteri
8	828.5	30.7	641	1 EBNI_EBV	P03211 epstein-bar
9	819.5	30.4	801	1 Y747_MYCTU	O53810 mycobacteri
10	807	29.9	957	1 Y278_MYCTU	P56877 mycobacteri
11	782.5	29.0	603	1 YD55_MYCTU	O10637 mycobacteri
12	736	27.3	384	1 GRP1_PETMY	P09789 petunia hyb
13	729.5	27.0	860	1 ELS_MOUSE	P54330 mus musculu
14	714.5	26.5	338	1 GRP_ARATH	P27483 arabidopsis
15	697.5	25.8	481	1 LORI_MOUSE	P18185 mus musculu
16	684	25.3	543	1 YP91_MYCTU	O50630 mycobacteri
17	680.5	25.2	864	1 ELS_RAT	O09372 rattus norv
18	671.5	24.9	498	1 Y418_MYCTU	O50615 mycobacteri
19	612	22.7	463	1 Y418_MYCTU	O50615 mycobacteri
20	609.5	22.6	515	1 Y140_MYCTU	O50594 mycobacteri
21	608.5	22.5	434	1 YK96_MYCTU	O10707 mycobacteri
22	595	22.0	747	1 ELS_BOVIN	P04985 bos taurus
23	593.5	22.0	750	1 ELS_CHICK	P07916 gallus gall
24	536.5	19.9	730	1 ELS_HUMAN	P15502 homo sapien
25	526.5	19.5	252	1 GRP1_PHAUV	P33485 phaeoculus v
26	516	19.1	1733	1 VNUA_PPVKA	P33485 pseudorabie
27	512.5	19.0	672	1 PHX5_MOUSE	P08389 mus musculu
28	470	17.4	2038	1 FSH_DROME	P13709 drosophila
29	455.5	16.9	966	1 FIB1_PETMA	P02674 petromyzon
30	454	16.8	316	1 LORI_HUMAN	P23490 homo sapien
31	439	16.1	671	1 CALI_RAT	P02454 rattus norv
32	434	16.1	1027	1 CAPF_RTEPA	P30754 rattus pach
33	432	16.0	1156	1 GLH4_CABEL	O76743 caenorhabdi

34	429.5	15.9	1453	1 CAL1_CHICK	P02457 gallus gall
35	428.5	15.9	1466	1 CAL3_HUMAN	P02461 homo sapien
36	428	15.9	1464	1 CAL1_HUMAN	P02452 homo sapien
37	427	15.8	1453	1 CAL1_MOUSE	P11087 mus musculu
38	422	15.6	1459	1 CAL2_MOUSE	P28481 mus musculu
39	420	15.6	1460	1 CAL1_CANFA	O9x517 canis famli
40	419	15.5	1373	1 CAL2_MOUSE	O01149 mus musculu
41	415.5	15.4	1366	1 CAL2_CANFA	O46392 canis famli
42	415	15.4	1418	1 CAL2_HUMAN	P02458 homo sapien
43	411	15.2	1049	1 CAL3_BOVIN	P04258 bos taurus
44	409.5	15.2	1567	1 ICEN_XANCT	P18127 xanthomonas
45	409	15.1	1366	1 CAL2_HUMAN	P08123 homo sapien

ALIGNMENTS

```

RESULT 1
SPDL_NEPCL STANDARD: PRT: 747 AA.
AC P19837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_Taxid=6915;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=90384959; Pubmed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 653-747 FROM N.A.
RX MEDLINE=94165036; Pubmed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
  proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
  Araneus bicentarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
CC -I- FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE
  CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
  PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
  WITH ELASTIC AMORPHOUS SEGMENTS.
CC -I- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37137; AAA29380.1; -
DR EMBL: U03848; AAB0212.1; -
DR PIR: A36068; A36068.
KW Silk; Repeat.
FT DOMAIN 1 1 655 25 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 1 1 25
FT REPEAT 2 2 38
FT REPEAT 3 3 66
FT REPEAT 4 4 96
FT REPEAT 5 5 130
FT REPEAT 6 6 158
FT REPEAT 7 7 191
FT REPEAT 8 8 204
FT REPEAT 9 9 235

```

FT	REPEAT	236	262	10.
FT	REPEAT	263	292	11.
FT	REPEAT	293	305	12.
FT	REPEAT	306	333	13.
FT	REPEAT	334	360	14.
FT	REPEAT	361	394	15.
FT	REPEAT	395	424	16.
FT	REPEAT	425	458	17.
FT	REPEAT	459	485	18.
FT	REPEAT	486	512	19.
FT	REPEAT	513	525	20.
FT	REPEAT	526	555	21.
FT	REPEAT	556	582	22.
FT	REPEAT	583	612	23.
FT	REPEAT	613	642	24.
FT	REPEAT	643	655	25.
FT	CONFLICT	662	662	V -> L (IN REF. 1).
FT	CONFLICT	672	672	S -> T (IN REF. 1).
FT	CONFLICT	695	747	NPGSGCDVLTOALLEVVSALIQILSSSISIGOVNGSAQA TOIYGOSVYQYL -> ILVFIDVMSSFFLFRLELLLSRS (IN REF. 1).
SEQ	SEQUENCE	747 AA;	60528 MW;	850E44BD0649E012 CMC64;
<hr/>				
Query Match 84.2%: Score 2273; DB 1; Length 747;				
Best Local Similarity 71.0%: Pred. NO. 1.le-95;				
Matches 494; Conservative 4; Mismatches 10; Indels 188; Gaps 15.				
OY	6	GGOQMGRIRIRINGYGGTGLGSOAGOGGAAAAAAGAGAAGGAGOGGYGGLGSQGAGRGGGAGAAA	65	
Db	22	GGGAGAG---GGYGLGGGAGAGAGGAAAAAAGAGAGOGGYGGLGSQGAGRGGGAGAAA	78	
OY	66	AAAGAGOGCGTGGLGSQGAGRGLGGGAG--AAAAAGVGCGGGLG---OGAGGA--CA	119	
Db	79	AAAGAGOGCGYGGTLGSOGAGRGGLGGGAGAAAAAAGAGAGGGYGGTGNAGAGRGGA	138	
OY	120	AAAAAGAGOGGYGGTGLGSOGAGRGGSGGGAGAAAAAAG-----GAGGAGYG	166	
Db	139	AAAAAGAGAGGGYGGTGLGSOGAGRGGLGGGAGAAAAAAGAGAGOGGTGGTGGAGAGGGYG	198	
OY	167	GLGSOGAGRGGLGGGAGAG-----	185	
Db	199	GLGSOGAGRGGLGGGAGAGAAAAAAGAGAGGGTGLGGGAGAGASAAAAAGAGGGYGGT	258	
OY	186	-----AAAAAAGAGAGGGYGGTGLGSOGAGOGGTGGLGSOGAGRGLGGGAGA	232	
Db	259	GSOGAGRGEGAGAAAAAAGAGAGGGYGGTGLGCAGAGOGGYGGLGSOGAGRGLGGGAG-	317	
OY	233	AAAAGAGCGGGLG----GGCAGAAAAAAGAGAGOGGYGGLGSOGAGR--GGGAGA--AA	284	
Db	318	AAAAGAGCGGGLGGGAGAGAAAAAAGAGAGGGYGGTGLGSOGAGRGLGGGAGAVMAA	377	
OY	285	AAGGAGOGGYG---OGAGG-----GYGTLGSOGAGRGLGGGAGA	324	
Db	378	AAGGAGOGGYGGLGSOGAGRGGGAGAAAAAAGAGAGQRGVGGTGNAGAGRGLGGGAGA	437	
OY	325	AA-----AAAAAGAGGQ-----	336	
Db	438	AAAAAAGAGAGGGYGGTGNAGAGRGGQAGAAAAAAGAGAGGGYGGTGLGSOGAGRGGGAGAA	497	
OY	337	-----GSLGGGAGAAAAAAGAGAGGGTGLGGGAG	365	
Db	498	AAAVGAGDEGIRMGAGAGGGYGGTGLGSOGSGRGGLGGGAGAAAAAAGAGAGGGTGLGGGAG	557	
OY	366	OGAGAAAAAAGAGVRGGYGGTGLGSOGAGRGGGAGAAAAAAGAGAGOGGTGGLGGGAG	425	
Db	558	OGAG-----AAAAAGGVROGCTGGTGLGSOGAGRGGGAGAAAAAAGAGAGGGYGGTGGGAG	613	
OY	426	GAGTGGGAGAAAAAAGAGOGGYGTVGVSASAASAAASRLSPQASSRVSSAVSNLVASG	485	
Db	614	GRGGLGGGAGAAAAAGAGAGGGYGGTVGVSASAASAAASRLSPQASSRVSSAVSNLVASG	673	
OY	486	PTNSAALSTISNVVQIGASNPLGSCDVLIALL	521	

[illegible]

OY	104	-GGGGLGCGAGAGCGACAAAAAGAAGCGGCTC-----GLCSQAGAGCGSGCGG-AGAAAAAA	157
Db	123	SGGGCGPGGYPGPQ-----OQGEGTTPGQQGPGGTGTPGPGQQGSPGSAAAAAAA	177
OY	158	GGAGQ----GGG-----GLGSQAGARGGLGCGGAGAAAAAAGAAGCGTGATGCGAGCG	210
Db	172	SGPQQDQGRGCGPPQQDQPCGCTGPPQQDQSPSGPSAAAAAASGFGQDQPGGCTGPGQQDQPG	233
OY	211	GYGGLGQSQAGRGLGCGAGAAAAAAGAACGCGLGCGAG-----AAAAAAGCAQ	261
Db	232	GY-----GPGQQGLSGPGSAAAAAAGPGQQGPGGTYCPGQQGSPSGSAAAAAAAGP	285
OY	262	CGTC-----GLCSQAG-RCGQAG-AAAAAAAAAGAGCGCTGGCGAGCGCTGLSQAGRG	315
Db	286	GGYPPGQQGPGCYPPGQQGPGSAAAAAAGPGQQGLGCTGPGQQDQPGG---YGGQQ	342
OY	316	GLGSGAGAAAAAAGAGAGCGCGGLGCGGAG-----AAAAAAG-----GAGCGG	358
Db	343	GPGGTGSGSASAAAAAAGPGQQGCGCTGPGQQGSPSGPSAAAAAAGAPGCTGPGQQG	402
OY	359	LGGGGAQ-----GAGAAAAAAGAVRGQGYGLGSGAG-----RGQCGAG-AAAA	407
Db	403	PGGYAPGQQGSPSGPSASAAAAAAG--PGGYCP-QQQGPGGAYAPGQQGSPSGPSAAA	459
OY	408	AAGGAGCGGCTC-----GLGCGCV-----GAGCTG-----GGCAGAAAVAGCGG	447
Db	460	AAAAAGCGGCTGPAQCGSPGPIASAASAGCGGTGPAQCGPAGCPGSAVAASAGSAG	519
OY	448	YGVGSGASASAAASFLSSPOASRRVSANVYAAGPTNMAALSTTNVYSQTASN	507
Db	520	YG-----PGSQASAAASHLASPDGARVASAVSNLYSGPTYSAAALSSVISNAVSQTASN	575
OY	508	PGLSGCDVLIDALL 521	
Db	576	PGLSGCDVLIDALL 589	
<hr/>			
RESULT 3			
FBOH_BOMMO			
ID FBOH_BOMMO STANDARD:		PRT: 5263 AA.	
AC P05790: Q26379: Q17220:			
DT 01-NOV-1988 (Rel. 09, Created)			
DT 01-OCT-2000 (Rel. 40, Last sequence update)			
DT 01-OCT-2000 (Rel. 40, Last annotation update)			
DE FIBROIN HEAVY CHAIN PRECURSOR (FIB-H) (H-FIBROIN).			
GN FIBH.			
OS Bombyx mori (Silk moth).			
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;			
OC Bombycoidea; Bombycidae; Bombyx.			
OX NCBI_TaxID=7091.			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=20330362; PubMed=10871375;			
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,			
RA Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;			
RT "Fine organization of Bombyx mori fibroin heavy chain gene.";			
RL Nucleic Acids Res. 28:2413-2419(2000).			
RN [2]			
RP SEQUENCE OF 1-168 FROM N.A.			
RX MEDLINE=80045039; PubMed=498286;			
RA Teujimoto Y., Suzuki Y.;			
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'			
RT flanking, mRNA coding, entire intervening and fibroin protein coding			
RT regions."			
RL Cell 18:591-600(1979).			
RN [3]			
RP PARTIAL SEQUENCE FROM N.A.			
RX MEDLINE=79211211; PubMed=455439;			
RA Teujimoto Y., Suzuki Y.;			
RT "Structural analysis of the fibroin gene at the 5' end and its			
RT surrounding regions."			
RL Cell 16:425-436(1979).			
RL			

[4] PARTIAL SEQUENCE FROM N.A.
RP STRAIN-KINSHU X SHOWA;
RX MEDLINE=89094868; PubMed-3210244;
RT Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RL structure of silk fibroin mRNA."
RN J. Mol. Biol. 203:917-925(1986).
[5]
RN PARTIAL SEQUENCE FROM N.A.
RP MEDLINE=94365842; PubMed-7916056;
RX Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RL gene";
RN J. Mol. Evol. 38:583-592(1994).
[6]
SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RP STRAIN-JF-139;
RC MEDLINE=99296390; PubMed-10366732;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RT Takagi T., Mizuno S.;
RT "Determination of the site of disulfide linkage between heavy and
RL light chains of silk fibroin produced by Bombyx mori.";
RL Biochim. Biophys. Acta 1432:92-103(1999).
CC -I- FODCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC -I- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC -I- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC SECTION OF SILK GLANDS.
CC -I- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC ALA-GLY-ALA)-N INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC AMORPHOUS REGIONS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF26688; AAF76983.1; -
DR EMBL; V00094; CAA23432.1; -
DR EMBL; V00097; CAA23433.1; -
DR EMBL; S74439; AAB31861.1; -
DR EMBL; X13869; CAA32076.1; -
DR EMBL; M35378; AAA27839.1; -
DR EMBL; AB017362; BAA33147.1; -
PIR: S01844; S01844.
KW PIR; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 5263
FT DOMAIN 149 5206
FT DISULFD 5244 5244
FT DISULFD 5260 5263
FT CONFLICT 10 10
SQ SEQUENCE 5263 AA; 391586 MW; 8EE1ID3A0A47440E CRC64;

C -> V (IN REF. 2).
HIGLY REPETITIVE.
INTERCHAIN (WITH LIGHT CHAIN).

POTENTIAL.
FIBROIN HEAVY CHAIN.

[illegible]

```

CC or send an email to license@slsb.sib.ch).
CC -----
DR EMBL: AL022022; CAA1745.1; -.
DR HSSP: P19972; 1KVD.
DR Tuberculist; Rv3508; -.
DR InterPro; IPR000084; -.
DR Pfam; PF00934; PE; 1.
DR Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA: 147627 MW: C7B1923D5D0146CD CRC64;

Query Match 34.6%; Score 935; DB 1; Length 1901;
Best Local Similarity 37.0%; Pred. No. 1,1e-35;
Matches 264; Conservative 24; Mismatches 220; Indels 206; Gaps 28;

OY 5 TGGG--QMGRTIRIRYGLGGGAGGAGAAA---AAGCAG-OGYTGGLSGGAGRCG 58
   |||
Db 663 TCGAAGSVGMAGIGGTGTGTGGVGAAGAGAAAAGSSATGCGAFGAGAGGEGAGAGNSGV 722
   |||
OY 59 OGAGAAAAAGAGAGGCGYGLSGSOGAGR-----GGLGGGAGAAAAAGVGCGGLSGQ 111
   |||
Db 723 GGTNSSGAGAGAGCGTCGTGAGAGSSADNPITGAFAGAGCTC-GAAGAGAGAGATGTGT 781
   |||
OY 112 GAGGAGAAAAAAGAGGCGYTGGLSGGAG-----RGSGGCGAGAAAAAAG----- 159
   |||
Db 782 GGVVATGTSAGTCGAGRCRGDGDPAASGLGLSLGFPDGGGCGGAGSAGAGCTINGAGG 841
   |||
OY 160 -AGGCGYGLSGDSOGAG-----RGGLGGG-GAGAAA-----AAAAGAGAGG 197
   |||
Db 842 AGGNGGDDGDTGTAAGAGLADNGGVGVDGAGAGAAAGCAGAGVGLTRAKAGDGAAGCAGG 901
   |||
OY 198 GYGGGLGGGQ---AGGCGYGLSGSQG-----AGRGGLGGG-GAGAAA 235
   |||
Db 902 GAGGAGGAGADNNFNGCGGAGGCGGCGGLGAGASTTINANGAGAGNGCGTGGKAGAGACT 961
   |||
OY 236 AGGAGCGGLGGGGA-----GAAAAAGGAGAGCGTGGGLSGGGA-----G 273
   |||
Db 962 LGVGGSGGTGCGDGGAGSGCGGCGFGGAAAGKAGGGGNGCGDGGDCAAGLGLSLGFPDGG 1021
   |||
OY 274 RCGGAGAGAAAAGGA-GGGGYCGGAGACGGYGGYGLSGSOGAG-RGGLGGG-GAGAAA--- 326
   |||
Db 1022 CGGGGAGAGSAGAGCTINGAGAGCAG-EDGEGGATGAAGLGDNGGVGGGAGGAGAAAGNGG 1080
   |||
OY 327 -----AAAAAGGA-----GGGGLGGGQ-----AGAAAA-- 350
   |||
Db 1081 NAGVGLTRAKAGDGAAGNGAGAGAGAGADNNFNGCGGAGGCGGCGGLGAGASTTISIN 1140
   |||
OY 351 -----AGGAGCGGGLGGGQ-----AGGAGAAAAAAAAGAGGROGG 386
   |||
Db 1141 ANGAGAGNGCTGKCGAGAGATLVGGSSGCTGCGDGDAGSGGGGCGAAGCAAGGGGNGG 1200
   |||
OY 387 YGGLSGGQA-----GRGGGAGAAAAAAGG-----AGGGGTGGLGGGQV----- 425
   |||
Db 1201 VGGDGEAGASGLGLSLGSPDGGGCGGAGGAGAGCTGTGAGAGDGAAPATLI 1260
   |||
OY 426 -----GAGLGGGAGAGAAAAGV-----GGGTYGVGSAGSASAA 460
   |||
Db 1261 GGPDDGGDGGGIGDGGAGAGCAAGPDDGGDGGAGNAGFAGVPGGGCTGTGAGAGAGGA 1320
   |||
OY 461 AASRLSPQASSRVSASVSNILVASPTNSAALSTITSINVSIGASNPGLSGCD 514
   |||
Db 1321 GADGPPSIDGGGAGAGCHGCGCGKGLNLTGLAASAGDGGAGAGAGGNGGD 1374

```

```

DE 01-OCT-1994 (Rel. 30, Last annotation update)
OS GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1.8 PRECURSOR (GRP 1.8).
OC Phaeosulfus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Phaseolus.
ON NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. TENDERGREEN;
RX MEDLINE=69091109; PubMed=3208742;
RA Keller B., Sauer N., Lamb C.J.;
RT "Glycine-rich cell wall proteins in bean: gene structure and
RL association of the protein with the vascular system.";
EMBL J.73625-3633(1988).
CC -1 FUNCTIONAL RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1 SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC -1 DOMAIN: THE REPEATED DOMAINS OF THE PROTEIN FORM A BETA-PLAIED
CC SHEET CONFIGURATION.
CC -1 SIMILARITY: THE N-TERMINAL SIGNAL SEQUENCE OF ABOUT 30 AA SHOWS
CC ABOUT 60% HOMOLOGY TO THAT OF THE GLYCINE-RICH CELL WALL
CC STRUCTURAL PROTEIN GRP 1.0.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isdb.ch/announce/,
CC or send an email to license@isdb-sib.ch).
CC -----
DR EMBL: X13596; CAA31932.1; -.
DR PIR: S01820; S01820.
DR HSSP: P30129; ADPV.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 465 GLYCINE-RICH CELL WALL STRUCTURAL
FT FT PROTEIN 1.8.
FT FT
FT DOMAIN 33 465 GLY-RICH.
FT FT DOMAIN 205 359 8 x 22 AA TANDEM REPEATS.
SQ SEQUENCE 465 AA; 36683 MW; B5C4A9B983BA3607 CRC64;
-----
Query Match 32.1%; Score 866.5; DB 1; Length 465;
Best Local Similarity 47.8%; Pred. NO. 5.1e-33;
Matches 218; Conservative 8; Mismatches 177; Indels 53; Gaps 19;
OY 23 GGGAGCGAGAAAAAAGGAGCGGCGYGGLSGQSGRGGCGAGAAAAAAGGAGCGGYSGLSGO 82
DB 33 GTGLCHGTGGTGCAAGSTVGCGGGGGGGGAGABEHGVGTGGSGGGGGGVTGGGX 92
OY 83 GAGRGGLGCGGAGAAAAGATGATGCGGAGAGAAAAGAGAGCGGATGSLGSQAGR 142
DB 93 GAGTGGGGGGGGGGGAVATGGGGE--RGVGGGGG-----GGAG-GGYAGNGHGTCT 141
OY 143 GCGSGGCGAGAAA-AAGGAGCGGCTGGLSGQAGKGC-----LGCGAGAAAAAACG 193
DB 142 GGGGGSGAGGGGGYNAAGGAAGCGGTGTGGGAGCGGGGGGHDHGSGYGGCGAGGAGGTYG 201
OY 194 AGGGYGGSLGGCGAGG--GGYGGSLGSGAGRGRLGGGCGAGAAAAGAGAGCGGLGCGAGAA 252
DB 202 GGEHHGGGGGGGGGAGGGGATGAGHGGGAG--GGGCGGAGGGGTAGCGNH--GAGGGGGG 258
OY 253 AAAAGAGACGGATGATGSGAGCGAGCGAGAAAAGAGAGCGGATGCGAGAGCGATGSLSGA 312
DB 259 GGAGGCTAGCGENH--GAGAGCGGG--GAGGCTAGCGEHHGGGAGGCGGAGGCTAGCGENH 316
OY 313 GRGSLGCGAGAAAAAAGAGAGCGGSLGGCGAGAAAAAAGAGAGGGL--GGGAGAGAGA 370
DB 317 GGGG--GGGCGGAGGGYAAVGHGGGGYGGGGG---DGGGCTGGEGHGGGYYGGGGG-- 370
OY 371 AAAAAAAAAAGCTGGGTGGLSGQGAGRGCGAGAAAAAAG--GAGGCGTGL----- 420

```

DB	Accession	Score	Length	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10	DB 11	DB 12	DB 13	DB 14	DB 15	DB 16	DB 17	DB 18	DB 19	DB 20	DB 21	DB 22	DB 23	DB 24	DB 25	DB 26	DB 27	DB 28	DB 29	DB 30	DB 31	DB 32	DB 33	DB 34	DB 35	DB 36	DB 37	DB 38	DB 39	DB 40	DB 41	DB 42	DB 43	DB 44	DB 45	DB 46	DB 47	DB 48	DB 49	DB 50	DB 51	DB 52	DB 53	DB 54	DB 55	DB 56	DB 57	DB 58	DB 59	DB 60	DB 61	DB 62	DB 63	DB 64	DB 65	DB 66	DB 67	DB 68	DB 69	DB 70	DB 71	DB 72	DB 73	DB 74	DB 75	DB 76	DB 77	DB 78	DB 79	DB 80	DB 81	DB 82	DB 83	DB 84	DB 85	DB 86	DB 87	DB 88	DB 89	DB 90	DB 91	DB 92	DB 93	DB 94	DB 95	DB 96	DB 97	DB 98	DB 99	DB 100	DB 101	DB 102	DB 103	DB 104	DB 105	DB 106	DB 107	DB 108	DB 109	DB 110	DB 111	DB 112	DB 113	DB 114	DB 115	DB 116	DB 117	DB 118	DB 119	DB 120	DB 121	DB 122	DB 123	DB 124	DB 125	DB 126	DB 127	DB 128	DB 129	DB 130	DB 131	DB 132	DB 133	DB 134	DB 135	DB 136	DB 137	DB 138	DB 139	DB 140	DB 141	DB 142	DB 143	DB 144	DB 145	DB 146	DB 147	DB 148	DB 149	DB 150	DB 151	DB 152	DB 153	DB 154	DB 155	DB 156	DB 157	DB 158	DB 159	DB 160	DB 161	DB 162	DB 163	DB 164	DB 165	DB 166	DB 167	DB 168	DB 169	DB 170	DB 171	DB 172	DB 173	DB 174	DB 175	DB 176	DB 177	DB 178	DB 179	DB 180	DB 181	DB 182	DB 183	DB 184	DB 185	DB 186	DB 187	DB 188	DB 189	DB 190	DB 191	DB 192	DB 193	DB 194	DB 195	DB 196	DB 197	DB 198	DB 199	DB 200	DB 201	DB 202	DB 203	DB 204	DB 205	DB 206	DB 207	DB 208	DB 209	DB 210	DB 211	DB 212	DB 213	DB 214	DB 215	DB 216	DB 217	DB 218	DB 219	DB 220	DB 221	DB 222	DB 223	DB 224	DB 225	DB 226	DB 227	DB 228	DB 229	DB 230	DB 231	DB 232	DB 233	DB 234	DB 235	DB 236	DB 237	DB 238	DB 239	DB 240	DB 241	DB 242	DB 243	DB 244	DB 245	DB 246	DB 247	DB 248	DB 249	DB 250	DB 251	DB 252	DB 253	DB 254	DB 255	DB 256	DB 257	DB 258	DB 259	DB 260	DB 261	DB 262	DB 263	DB 264	DB 265	DB 266	DB 267	DB 268	DB 269	DB 270	DB 271	DB 272	DB 273	DB 274	DB 275	DB 276	DB 277	DB 278	DB 279	DB 280	DB 281	DB 282	DB 283	DB 284	DB 285	DB 286	DB 287	DB 288	DB 289	DB 290	DB 291	DB 292	DB 293	DB 294	DB 295	DB 296	DB 297	DB 298	DB 299	DB 300	DB 301	DB 302	DB 303	DB 304	DB 305	DB 306	DB 307	DB 308	DB 309	DB 310	DB 311	DB 312	DB 313	DB 314	DB 315	DB 316	DB 317	DB 318	DB 319	DB 320	DB 321	DB 322	DB 323	DB 324	DB 325	DB 326	DB 327	DB 328	DB 329	DB 330	DB 331	DB 332	DB 333	DB 334	DB 335	DB 336	DB 337	DB 338	DB 339	DB 340	DB 341	DB 342	DB 343	DB 344	DB 345	DB 346	DB 347	DB 348	DB 349	DB 350	DB 351	DB 352	DB 353	DB 354	DB 355	DB 356	DB 357	DB 358	DB 359	DB 360	DB 361	DB 362	DB 363	DB 364	DB 365	DB 366	DB 367	DB 368	DB 369	DB 370	DB 371	DB 372	DB 373	DB 374	DB 375	DB 376	DB 377	DB 378	DB 379	DB 380	DB 381	DB 382	DB 383	DB 384	DB 385	DB 386	DB 387	DB 388	DB 389	DB 390	DB 391	DB 392	DB 393	DB 394	DB 395	DB 396	DB 397	DB 398	DB 399	DB 400	DB 401	DB 402	DB 403	DB 404	DB 405	DB 406	DB 407	DB 408	DB 409	DB 410	DB 411	DB 412	DB 413	DB 414	DB 415	DB 416	DB 417
----	-----------	-------	--------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

```

Db      NPNTGANTDMSGNNGNQNGNGRGPRAGVGEABGV--GGQGGIGLESJLDGNDGTGKGGA   444
OY      184 GAAAAAAGAAGAGCGGYGLG-----GCGAGCGCYGCLGSQC-----AGRGGLAGGAGGA   232
         445 GGTAAGTDDGACGAGAGAGGIGETDGSAGAGVAATGEBGGDGTATGVDGAGGAGKGGQGHT   504
OY      233 AAAAGNAGGGGLGCGAGAAAAMAAAGNAGCGGVCGLGSQSG-----AGRGGCAG   280
         505 GVGDAFEGDGGTGIGSGNGLALAAAGNGGTGAGAGGNGGGMILGNGAGAGAGTGTCTGG   564
OY      281 AAAAAAGAGAGCGGYGCGSQ-----AGCGYGGLGSSQC-----AGRGGLGCGAGAAA   326
         565 CAGACGAGV--CGAGGEBLTDGACIAGEBTGGLGGLGCVGTGGMGSSGGYVGNGGAAGS   622
Db      623 LIGGGGGGAGAGVGGTG-----GIGGIGAGAGNGGAGAGTTGGGATIGGGGTGG   674
OY      327 AAAAAGAGAGCGGLGCGGAGAAAAAAGAGAGCGGLGCGGAGGAGAAAAAAGAGVRGG   386
         623 LIGGGGGGAGAGVGGTG-----GIGGIGAGAGNGGAGAGTTGGGATIGGGGTGG   674
OY      387 YGGLGASGC--AGRGSC--GAGAA-----AAAAGAGCGGT---GGLGGCGVGAAGLGGQ   433
         675 VGGAGGTGTGTGAGAGTGTGGAGAGGLICWAGMAAGTGTAAGCGGGLGCGG--GNGNGGT   733
OY      434 GA----GAAAVGAGCGGCGYGVGSGASAASAASTLSP   468
         734 GATGGGCGDFALG--GNGGAGGAGGSGSGSSGTGCMGPP   771

RESULT       7
WA22_MYCTU   STANDARD:          PRT:    914 AA.
ID           WA22_MYCTU
AC           006794;
DT           30-MAY-2000 (Rel. 39, Created)
DT           30-MAY-2000 (Rel. 39, Last sequence update)
DE           30-MAY-2000 (Rel. 39, Last annotation update)
DE           WAG22 ANTIGEN PRECURSOR.
GN           WAG22 OR RV1759C OR MTCY28_25C.
OS           Mycobacterium tuberculosis.
OC           Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC           Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX           NCBI_TaxID:1773;
RN           [1]
RC           SEQUENCE FROM N.A.
RP           STRAIN=H37RV;
RX           MEDLINE=98295987; PubMed=9634230;
RA           COLE S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA           Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA           Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA           Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S.,
RA           Hovnsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA           Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,
RA           Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA           Taylor K., Whitehead S., Barrall B.G.;
RT           "Deciphering the biology of Mycobacterium tuberculosis from the
RT           complete genome sequence.";
RL           Nature 393:537-544(1998).
CC           -1 SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC           SUBFAMILY.
CC           -----
CC           This SWISS-PROT entry is copyright. It is produced through a collaboration
CC           between the Swiss Institute of Bioinformatics and the EMBL Outstation --
CC           the European Bioinformatics Institute. There are no restrictions on ways
CC           use by non-profit institutions as long as its content is in no way
CC           modified and this statement is not removed. Usage by and for commercial
CC           entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC           or send an email to license@isb-sib.ch).
CC           -----
DR      EMBL: Z95890; CAB09322.1; -.
DR      HSSP: P41140; 2SFA.
DR      Tuberculist; RV1759c; -.
DR      InterPro: IPR000084; -.
DR      Pfam; PF00934; pe.1.
KW      Antigen; Repeat; Signal.
```

FT SIGNAL	1	30	POTENTIAL.
FT CHAIN	31	914	WAG21 ANTIGEN.
SO SEQUENCE	914 AA;	74354 MM;	F6953C3DBEB86AC8 CRC64;
Query Match	31.1%;	Score 839.5;	DB 1; Length 914;
Best Local Similarity	37.4%;	Pred. No. 1.3e-31;	
Matches 249;	Conservative 22;	Mismatches 210;	Indels 185; Gaps 29;
OY	6	GGQMGIRIRIRINGYGLGCGAGACGAGAAAAA-----AGAGG-----GGYGLG	50
DB	129	GGDAGMLTGNCGAGSSGAGKANGAGGPGGAAGLFGNGAGAGAGGTATANNIGGAGAG	188
OY	51	SO-----GAGRGGCGAGAAAAAG-----AGGGYGLGSGGAGRGGLG-----QGAGA	96
DB	189	GSAMLEFAGAGAGAGAAATSLVGCIGTGTGTGNAGMLAGAAAGAGAGAGGFSEFSTAGAG	248
OY	97	AAAAGV-----GGGGLGCGAGAGCGAGAAAAA-----GGAAGGGYGLGSGQGA-GRGSGGQ	148
DB	249	AGGAGGLTTTGGVGGAGGCGGHTCGAGAGAGAGLFGAGAGMGAGGCFDGHGTGCTGAGAGD	308
OY	149	GAGAAAAAAGAGAGGGYGLGSGGAGRGGLGCGGCGAGAAAAAAGAG-----GG	198
DB	309	GGGGGLFAGAG--GDDGAGGSGLTGTGAAGNGG--NAGTLSTGAAGAGAGGTGAGAGTVPFG	365
OY	199	YGLGCGCA-----GGCGYGLGSGGAGRGGLGCGGCGAGAAAAAGAGCGG--AG	250
DB	366	KGAGAGAGAGNAGMLFGSGGGGCGTGGFGRPAAGGQGGVGGAGMLSSGSGGAGGSGGPG	425
OY	251	AAAAAAGA-----GGGGYGLGSGGAGRGG--GAGAAAAAAGAGAGGGYGGG--AG	300
DB	426	TAAAGAGAGAGAPGLIGNGNGNGGSGEGTGGVGGAGAGNAVLTGNGBEGTIGALAKSG	485
OY	301	GGYGGGL-----GGGAG--R	314
DB	486	FGGFGGLTADGYNAPESTSPWHNLQDDILSFINEPTEALTRPLIGNGDSGTPTGDD	545
OY	315	GGGCG-----GCGAGAAAAAAGAG-----GGGGLGCGGAGAAAAAAG	352
DB	546	GGAGGMLFGNGGNGCAGAAATGNSAGGAGAGAGLFTGTGAGAGAGAGVGTAGAGGAGG	605
OY	353	GA-----GGGGLGCGGAGG-----GAGAAAAAAGAG-----GYNAGGYGLG	391
DB	606	SAFLIGSGGTGGVGAATTTGGVGAAGNGLTGAAGLGGCGGCAFTAGVTTTGGAGGTG	665
OY	392	-----SOGAGCGCGAGAAAAAAGAG-----GCGTGGGLGCGG--VGAGGLGCGA-	435
DB	666	GAGGFLFANCGAGAGAGCTGTAGAGAGAGAGGLYAHGCTGGCGAGGCTGAGAGTGAAGP	725
OY	436	-----GAAAAVAG-----GGGYGGGCGASAAASRLSPQASRYSSAVSNLYASGP	486
DB	726	GGLYAGAGSGGAGGAGCGMAAGCGGCGVGNAGSLTLNAAGAGGSGSSSLSGKAG--AGGA	782
OY	487	TNSAAL	492
DB	783	GGSAGL	788
RESULT	8		
EBN1_EBV			
ID	EBN1_EBV	STANDARD:	PRT: 641 AA.
AC	P03211;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	EBNA-1 NUCLEAR PROTEIN.		
GN	BKRF1.		
OS	Epstein-barr virus (strain B95-8) (Human herpesvirus 4).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
CC	Gammaherpesvirinae; Lymphocryptovirus.		
OX	NCBI_TaxID=10377;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

RX MEDLINE=84270667; PubMed=6087149.
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tufnell P.S., Bartell B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211(1984).
 RN [2]
 RP SEQUENCE OF 1-26 FROM N.A.
 RX MEDLINE=86259739; PubMed=3460083;
 RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
 RT "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
 RT proteins: a probable transcriptional initiation site.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=90266473; PubMed=2161150;
 RA Petri L., Sample C., Kieff E.;
 RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
 RT latent infection nuclear proteins.";
 RL Virology 176:563-574(1990).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
 RX MEDLINE=96006523; PubMed=7553871;
 RA Bochkarev A., Bartwell J.A., Plutner R.A., Furey W.F. Jr.,
 RA Edwards A.M., Freppier L.;
 RT "Crystal structure of the DNA-binding domain of the Epstein-Barr
 RT virus origin-binding protein EBNA 1.";
 RL Cell 83:39-46(1995).
 CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
 CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSCRIPTVATING FACTOR
 CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT
 CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
 CC WITH THE NUCLEAR MATRIX.

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.ebi.ac.uk/termsandconditions) or send an email to license@ebi.ac.uk.  
CC -----  
DR EMBL; V01555; CAZ4816.f; -.  
DR EMBL; M13941; AAA45889.1; -.  
DR PIR; A03773; OOBK31.  
DR PIR; S33021; S33021.  
DR PDB; 1VHT; 23-DEC-96.  
DR TRANSFAC; T00211; -.  
KW Nucleic protein; DNA-binding; Transcription regulation; Activator;  
FT 3D-structure.  
MD DOMAIN 87 352 GLY/ALA-RICH  
SV SEQUENCE 641 AA; 56427 MW; 4D161653E16FC341 CRC64;
```

Query Match 30.7%; Score 828.5; DB 1; Length 641;
Best Local Similarity 46.0%; Pred. No. 3..1e-31;
Matches 209; Conservative 18; Mismatches 152; Indels 75; Gaps 17;

Qy	40	GAGGGGCGGLSGOGA	GRGCGGA	AAAAAGA	GGOGGYG	LGSOGAR	GCGLSOGAGAA	A	99
Dd	5	GPGRPNGSLERKOTSPREG-----	SGSSPGRRRG-DNHGRGRGRGRGRCRGA	56					
Qy	100	AGGVQG-----	GLGGO--GAGO	GACA	AAAAGAGAGOGGYG	LSOGCA	140		
Dd	57	PGSSSGSRHNDGVRRPKRPSCTICKTHGTGAGAGAGAGAGAGAGAGAGAGAG	116						
Qy	141	GRGGGCGGAGAAAAAGAGAGOGGTGGLG	SOGARGGGLGSGOAGAAAAAGAGOGGTG	200					
Dd	117	GAGCAGGAGAG-----GAGAGG-----	GAGCAGGAGAG-----GAGAG--G	152					
Qy	201	GLGGGAGOGGVTGLGSGAGARGCLGSGO	GAGAAAAAGAGOGGYGLGSGOAGAAAAAGAG	260					

[illegible]

```

RESULT 9
v747_MYCNU STANDARD: PRT: 801 AA.
ID v747_MYCTU
AC 053610;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV0747 PRECURSOR.
OS RV0747 OR MTW041.21.
GN Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sultson J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AL021958; CAA17514.1; -.
CC DR TubercuList; RV0747; -.
CC DR InterPro; IPR000084; -.
CC DR Pfam; PF00934; Pf: 1.
CC KW Hypothetical protein; Repeat; Signal.
CC FT SIGNAL 1 30
CC FT CHAIN 31 801 POTENTIAL.
CC FT SIGNAL 1 801 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
CC FT RV0747.
CC SEQUENCE 801 AA: 65407 MW: 85454CBFA5A00F41 CRC64:

```



```

OY 17 GYGLGCGGAGAGAGAAAAAGCA-GGGYGLGSGQ---GAGRGGGAGAAAAAGGAG 72
    |||||
DB 137 GNGGAGGGGSGA--PGAGAGNGGAGAGLFGSGGAGAGSTDAVAGAGGAGGAGNAGMLFGAAG 195
OY 73 GGGYGLGSGGCA--GRGLGSGGCA--GAAAAAGVGOGGL--GGGAGGAGAAAAAGG 126
    |||||
DB 196 VGGVGGFBNNGATGAGAGAGAGLFGAGRERGGSGGNTLGGAGGAGAGNACTLATGTGGG 255
OY 127 AG-----GGYGLGSGCA-----GSGSGGCG-----AGAAAAAG----- 158
    |||||
DB 256 AGGTGAGSRSGGFGGAGAGGAGGAGMFSGSGGAGAGISKSVSDAAGAGAGAPGLIENG 315
OY 159 -----GAGGCGYGLGSGQ-----AGRGLGSGCA----- 183
    |||||
DB 316 GNGGNGGASTGGGCGGAGGTGVLIGNGGSGGTGATLKGAGIGGTGVLGLDGF 375
OY 184 -----GAAAAAGAGAG----- 195
    |||||
DB 376 TAPASTPLHTLQDQVIMVNDPFTLTGRPLIGNANGTPTGTADGAGGWLFGNGNG 435
OY 196 -----GGYGLGSGCA-----GGGCGYGLGSGCA--GRGLGCGGAGAAAAAGAGGCGG 243
    |||||
DB 436 GCGTIGVNVGAGAGAGAGLFGTGTGGSGGPGATGLGIGGAG--GALLLFGSGGAGG 494
OY 244 LGGGCA-----GAAAAAGAGAGCGGTGGLGSGGAGRGGCGGAGAAAAAGAGAGCGGTG 295
    |||||
DB 495 SGGAGAVGNGGAGAGNAGALLGAAAGAGGAGAVG--GNGGAGNGGLFANGAGGPGGF 553
OY 296 GCGAGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 349
    |||||
DB 554 GSPAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 613
OY 350 AAG-GAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402
    |||||
DB 614 AAGVSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 671
OY 403 GAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
    |||||
DB 672 SGGFSGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 731
OY 459 SAASRSLSPQASRVSASVSNLVASGPTNSAALSTISNVVSOIGASNPGLSGCDVLIQ 518
    |||||
DB 732 AGGATGTLGSAFHTGGKGVGGSVAVLIGNGNGNGNGNSGNAGKSSGA--PGFSGAGAGAG 789
OY 519 ALLGHR 524
    |||||
DB 790 LLLGEN 795

```

```

RA Taylor K., Whitehead S., Barrell B.G.:
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
CC Nature 393:537-544(1998).
CC -I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGMS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021930; CAAL7353.1; -.
DR TubercuList; RV0278C; -.
DR InterPro; IPR000084; -.
DR Pfam; PF00934; PE; 1.
DR Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 957
FT POTENTIAL.
FT HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV0278C.
SQ SEQUENCE 957 AA; 81905 MM; 71EBABDA17FBA47C CRC64;

Query Match 29.9%; Score 807; DB 1; Length 957;
Best Local Similarity 36.0%; Pred. No. 3.7e-30;
Matches 219; Conservative 21; Mismatches 195; Indels 174; Gaps 22;

OY 17 GYGLGCGGAGAGAGAAAAAG--GAGGCGYGLGSGGAGGCGGAGAAAAAGAGCG 74
    |||||
DB 265 GYGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 323
OY 75 GYGLGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
    |||||
DB 324 GSGGSHVAVGCDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 381
OY 124 AGGAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 173
    |||||
DB 382 FSGSGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 440
OY 174 GRGGLGCGGAGAGAAAAAG--AGGCGYGLGSGGAGCGGAGAGAGAGAGAGAGAG 215
    |||||
DB 441 GNGGNGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 500
OY 216 -----GSGAGR--GGLG----- 227
    |||||
DB 501 APASTPLHTLQDQVIMVNDPFTLTGRPLIGNANGTPTGTADGAGGWLFGNANGT 560
OY 228 GAGAGAAAAAG--AGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 276
    |||||
DB 561 PGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 620
OY 277 -----GAGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 326
    |||||
DB 621 ICGAGAGAGVSLTISGGTGGNGNSIGVAGIGAGAGRGDAGLLFGAAGTGGRHAGAGV 680
OY 327 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
    |||||
DB 681 PAGVAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 740
OY 365 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 413
    |||||
DB 741 LFGAGGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 800
OY 414 -----GCGTGLGAGG-----YGAGLGGGAGAGAGAGAGAGAGAGAGAGAG 446
    |||||
DB 801 ALLFGRGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 860
OY 447 GYGGGSGA 455
    |||||
DB 861 GTGGDGGHA 869

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:05:35 ; Search time 95.41 Seconds
(Without alignments)
732.177 Million cell updates/sec

Title: US-09-490-291-8
Profile score: 2700

Sequence: 1 MASMTGGGQMGRIIRRGYGG.....GLSGCDVLLQALLGHHNNHH 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhcc:*
8: SP-organella:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP_unclassified:*
13: SP-vertebrate:*
14: SP_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2555	94.6	544	5	046171	046171 nephila cla
2	2313	85.7	617	5	046172	046172 nephila cla
3	1397	51.7	988	5	017434	017434 nephila cla
4	1099	40.7	636	5	016987	016987 araneus dia
5	968.5	35.9	2639	5	076786	076786 araneus dia
6	935	34.6	1884	5	09NMW2	09NMW2 nephila mad
7	927	34.3	1079	2	053557	053557 mycobacteri
8	921	34.1	871	5	044358	044358 nephila cla
9	912.5	33.8	1489	5	053559	053559 mycobacteri
10	900	33.3	2249	5	09NMW4	09NMW4 nephila cla
11	892	33.0	1381	2	053552	053552 mycobacteri
12	890.5	33.0	738	5	002402	002402 pinnecta fu
13	880	32.6	767	2	053435	053435 mycobacteri
14	877.5	32.5	1468	5	09GUB5	09GUB5 gallieria me
15	868.5	32.2	1660	2	053215	053215 mycobacteri
16	862	31.9	907	5	044359	044359 nephila cla
17	858	31.8	853	2	053439	053439 mycobacteri
18	857.5	31.8	882	2	053845	053845 mycobacteri
19	850	31.5	749	2	053844	053844 mycobacteri

20	847	31.4	626	5	09NMW1	09NMW1 nephila mad
21	847	31.4	1329	2	006810	006810 mycobacteri
22	843.5	31.2	714	2	053556	053556 mycobacteri
23	836	31.0	1538	2	053395	053395 mycobacteri
24	833	30.9	923	2	053890	053890 mycobacteri
25	831	30.8	1306	2	053775	053775 mycobacteri
26	827.5	30.6	741	2	006808	006808 mycobacteri
27	820.5	30.4	484	2	053394	053394 mycobacteri
28	818	30.3	667	2	053415	053415 mycobacteri
29	813	30.1	1011	2	P94986	P94986 mycobacteri
30	812.5	30.1	783	2	053809	053809 mycobacteri
31	810.5	30.0	610	5	Q9V5V8	Q9V5V8 drosophila
32	810	30.0	731	2	050415	050415 mycobacteri
33	801.5	29.7	410	5	016988	016988 araneus dia
34	796.5	29.5	837	2	053684	053684 mycobacteri
35	793	29.4	576	2	P71664	P71664 mycobacteri
36	783.5	29.0	408	10	043688	043688 vigna ungu
37	781.5	28.9	694	2	053212	053212 mycobacteri
38	775.5	28.7	390	10	09M3Y2	09M3Y2 triticum ae
39	767	28.4	294	5	016986	016986 araneus dia
40	764.5	28.3	606	2	053884	053884 mycobacteri
41	758.5	28.1	1729	5	Q9U617	Q9U617 drosophila
42	748.5	27.7	584	2	053575	053575 mycobacteri
43	747.5	27.7	618	2	006801	006801 mycobacteri
44	747	27.7	615	2	005806	005806 mycobacteri
45	742.5	27.5	639	2	053952	053952 mycobacteri

ALIGNMENTS

RESULT 1
ID 046171 PRELIMINARY; PRT; 544 AA.
AC 046171;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SPIDROIN 1 (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 449-544 FROM N.A.
RX MEDLINE=98148687; PubMed=9487707;
RA Arcidiacono S., Mello C., Kaplan D., Cheley S., Bayley H.;
RT "Purification and characterization of recombinant spider silk
RT expressed in Escherichia coli.";
RL Appl. Microbiol. Biotechnol. 49:31-38(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX Beckwith R., Arcidiacono S., Store R.;
RT Insect Biochem. Mol. Biol. 0:0-0(1998).
RN [4]
RP SEQUENCE FROM N.A.
RX Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20329; AAC38957.1; -;
FT NON-TER 1 1
SQ SEQUENCE 544 AA; 44107 MW; CC611B3551945615 CRC64;

Query Match 94.6%; Score 2555; DB 5; Length 544;
Best Local Similarity 99.6%; Pred. No. 5,86-143;
Matches 503; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy	17	GYGGLGGCGAGGAGAAAAAAGAGAGCGGTGGTGSOGAGRCGGCAGAAAAAGACGGG	76
Dd	1	GYGGLGGCGAGGAGAAAAAAGAGCGGTGGTGSOGAGRCGGCAGAAAAAGACGGG	60
Oy	77	GGTSGSAGRCRGTLGGCGAGAAAAGVGGGTGGTGAGCGAGAAAAAGAAGCGGTG	136
Dd	61	GGTSGSAGRCRGTLGGCGAGAAAAGVGGGTGGTGAGCGAGAAAAAGAAGCGGTG	120
Oy	137	SOGAGRCGSGGGCAGAAAAAACGAGCGGTGTLGSGAGRCGTGGCGAGAAAAAGAGCG	196
Dd	121	SOGAGRCGSGGGCAGAAAAAACGAGCGGTGTLGSGAGRCGTGGCGAGAAAAAGAGCG	180
Oy	197	GGYGGTGGCGAGGGGCTGGTGSOGAGRCGTGGCGAGAAAAAGAGCGGTGGCGAGAAAAA	256
Dd	181	GGYGGTGGCGAGGGGCTGGTGSOGAGRCGTGGCGAGAAAAAGAGCGGTGGCGAGAAAAA	240
Oy	257	GGAGCGGTGGTGSOGAGRCGGAGAAAAAGAGCGGTGGCGAGCGGTGGTGSOGAGRCG	316
Dd	241	GGAGCGGTGGTGSOGAGRCGGAGAAAAAGAGCGGTGGCGAGCGGTGGTGSOGAGRCG	300
Oy	317	LGGCGAGAAAAAAAGACGAGCGGTGGCGAGAAAAAGAGCGGTGGCGAGCGAGAAAAA	376
Dd	301	LGGCGAGAAAAAAAGACGAGCGGTGGCGAGAAAAAGAGCGGTGGCGAGCGAGAAAAA	360
Oy	377	AAAGCVGGGTGGTGSOGAGRCGGAGAAAAAGAGCGGTGGTGGCGGVGAGTGGCGAG	436
Dd	361	AAAGCVGGGTGGTGSOGAGRCGGAGAAAAAGAGCGGTGGTGGCGGVGAGTGGCGAG	420
Oy	437	AAAAVGAGCGGTGGVGSASASAARSLSPOASSRVSAVSNTLVASGPINSAALSSTI	496
Dd	421	AAAAVGAGCGGTGGVGSASASAARSLSPOASSRVSAVSNTLVASGPINSAALSSTI	480
Oy	497	SNVYSQTGAENPGLSCGDVLIOALL 521	
Dd	481	SNVYSQTGAENPGLSCGDVLIOALL 505	
<hr/>			
RESULT	2		
O46172	ID	PRELIMINARY:	PRT: 617 AA.
AC	O46172;		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DR	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	DRAGLINE SILK PROTEIN SPIDROIN 1 (FRAGMENT).		
OS	Nephila clavipes (orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
CC	Araconemorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.		
NCBI	Taxid=6915;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=94165058; PubMed=8120021;		
RX	Beckwith R., Arcidiacono S.;		
RT	"Sequence conservation in the C-terminal region of spider silk		
RT	proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus		
RT	bicentennarius (Araneidae).";		
RL	J. Biol. Chem. 269:6661-6663(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Beckwith R., Arcidiacono S., Stote R.;		
RL	Insect Biochem. Mol. Biol. 0:0-0(1998).		
DR	EMBL: U37520; AAC04504.1; "		
FT	NON_TER 1 617 1		
FT	NON_TER 617 617		
SEQ	SEQUENCE 617 AA: 49410 MW: 1EFQCE9269A832E2 CR664;		

	Query Match	85.7%	Score 2313;	DB 5;	Length 617;
	Best Local Similarity	80.1%	Pred. No. 1e-128;		
	Matches 482;	Conservative 1;	Mismatches 5;	Indels 114;	Gaps 7;
07	34	AAAAAGCAGCGCGTGGTGGCAGAGGCGAGAAAAAAGCAGCGGTGGTGGCAGAGRGGTGGCGG	93		

Db	1	AAAAAGAGCGCGTGCGLGSGGAGRGGGGAGAAAAAAGAGCGGCGTGLGSGGAGRGGLGCGG	60
QY	94	AG--AAAAAGVGCGGLGSGGAGCGAGAAAAAAGAGCGCGTGGTGLGSGGAGRGSGGGGAG	151
Db	61	AGAAAAAAGCGVCGGGLGSGGAGCGAGAAAAAAGAGCGCGTGGTGLGSGGAGRGSGGGGAG	120
QY	152	AAAAAAGAGCGCGTGCGLGSGGAGRGGLGSGGAGAAAAAAGAGCGCGTGGTGLGSGGAGCGG	211
Db	121	AAAAAAGAGCGCGTGCGLGSGGAGRGGLGSGGAGAAAAAAGAGCGCGTGGTGLGSGGAGCGG	180
QY	212	YGLGSGGAGRGGLGSGGAG--AAAAAAGAGCGGLG--GCGAGAAAAAAGAGCGCGT	265
Db	181	YGLGSGGAGRGGLGSGGAGAAAAAAGAGCGGLGCGAGAGAAAAAAGAGCGCGT	240
QY	266	GLGSGGAGRGCGCGAGAAAAAAGAGCGCGTGGCGAGCGCGTGLGSGGAGRGGLGSGGAGAA	325
Db	241	GLGSGGAGRGCGAGAAAAAAGAGCGCGTGGCGAGCGCGTGLGSGGAGRGGLGSGGAGAA	300
QY	326	AAAAAAG-----AAAAAAG-----	332
Db	301	AAAAAAGGAGCGGLGSGGAGCGAGAAAAAAGAGCGCGTGCGLGNGAGRGCGGAAAAAAGGA	360
QY	333	-----GAGCGGLGCGGAGAAAAAAGAGCG--GGLGCGGA-----	364
Db	361	GCGGCGGLGSGGAGRGGLGSGGAGAAAAAAGAGCGCGTGGTGLGSGGAGCGGCGTGLGSGG	420
QY	365	-----GCGAGAAAA-----AAAAAAGGVRGCGGTGGLGSGGAGRG	399
Db	421	RCGLGCGGAGAAAAAAGAGCGCGGLGCGGAGCGAGAAAAAAGGVRGCGGTGGLGSGGAGRG	480
QY	400	CGAGAAAAAAGAGCGCGTGCGLGSGGAGRGGLGSGGAGAAAAAAGAGCGCGTGGTGLGSGGASAS	459
Db	481	CGAGAAAAAAGAGCGCGTGCGLGSGGAGRGGLGSGGAGAAAAAAGAGCGCGTGGTGLGSGGASAS	540
QY	460	AAASRLSPQASRRVSSAVSNLVAASGPTNSAALSTTSNVSQIGANPGLSGCDVLIQA	519
Db	541	AAASRLSPQASRRVSSAVSNLVAASGPTNSAALSTTSNVSQIGANPGLSGCDVLIQA	600
QY	520	LL 521	
Db	601	LL 602	

RESULT	3		
ID	017434	PRELIMINARY;	PRT; 988 AA.
AC	017434.		
DT	01-JAN-1998	(TREMBLrel. 05, Created)	
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)	
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)	
DE	MINOR AMPULPATE SILK PROTEIN MISP1 (FRAGMENT).		
OS	Nephila clavipes (Orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.		
OX	NCBI_TaxID=6915;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98200471; PubMed=9541398;		
RA	Colgin M.A., Lewis R.V.;		
RT	"Spider minor ampulpate silk proteins contain new repetitive sequences		
RT	and highly conserved non-silk-like 'spacer regions'."		
RL	Protein Sci. 7:667-672(1998).		
DR	EMBL; AF027735; AAC14589.1; '-'		
DR	InterPro; IPR000817; '-'		
DR	PRIMS; PR00341; PRON.		
FT	NON_TER 1		
SQ	SEQUENCE 988 AA: 79082 MW; 461E03DF53F7085D CRC64;		

Query Match	51.7%;	Score 1397;	DB 5;	Length 988;
Best Local Similarity	47.7%;	Pred. No. 7.8e-75;		
Matches 345;	Conservative 23;	Mismatches 144;	Indels 212;	Gaps 22;

01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE PGRS-FAMILY PROTEIN (FRAGMENT).
 GN RV3512 OR MTW023.19.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
 NCBI_TaxId=1773;
 RX MEDLINE=98295987; PubMed=9634230;
 RC STRAIN=H37RV;
 RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 Hornsby T., Jags K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL022022; CAA17749.1; .
 DR TuberculList: RV3512; .
 DR InterPro: IPR002202; .
 DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
 FT NON_TER 1
 SO SEQUENCE 1079 AA; 81163 MW; A79718CCDB7AB97D CRC64;

Query Match 34.3%; Score 927; DB 2; Length 1079;
 Best Local Similarity 40.2%; Pred. No. 3.1e-47;
 Matches 238; Conservative 14; Mismatches 202; Indels 138; Gaps 23;

2 ASMTGGGQCGRRIRINGYGLGCGG-----AGCGAGAAAAAGAGCGGCGY----- 47
 DB 315 AGGAAGSNGTGTGANGTGGDGGGGAAGATAGSNGAGTGSAGNGTGGGSGGAGG 374
 48 -GLGSGAGCGGAGAGAAAAAGAG-----GCGYGGGLGSGG-----AGRGGLGCGA 94
 DB 375 DGTGGVGGKGGKNGADBEVGGAGAGSGSPNTSPGNGGCGGCGGAGAGAGAGG 434
 95 GAAAAAGVGGGGLGCGGAGAGAAAAAGAGCGGCGGLGSGAGAGSGGCG--GAGAA 153
 DB 435 GANGTAGNCGCGGAGCTG---GAGAAASATNG--GSGGAGGTGGD--GSGGAGGTGAGGT 489
 154 AAAAGAGCGGTGGGLGSGAGAGAAAAAGAGCGGCGGLGSGAGAGSGGCG--GAGAA 199
 DB 490 GGAAGDCCGGCGGAGGAGGAGGAGAGTGGNGGNTGTAGTGAAGAGNGAAKGGGA 549
 200 GGLGCGAGCGGCGGLGSGG-----AGRGGLGCGG-----A 230
 DB 550 GCGGCGGTGGTGGCGGAGCGGAGCTGCDTGTGGGCTVPAGSGGCGGAGGAGGCGGAGD 609
 231 GAAAAAGAGCGGLG-----GCGAGAAAAAGAGAGCGGCGGLGSGAGAGCGG--GAGA 281
 DB 610 GCGGCGGADGTGGNGNGNGNRNSNGNTGGAGGCGGAGAGAGAGAGSGGCGTGGNGAGG 669
 282 AAAAAAG-----GAGCGGTGGCG--AGCGGCG--GLGSGAGAGRGGLGCGGAGAAAAA 331
 DB 670 DAGDAGNGNGNGTGGGNGGNGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
 332 GAGAGGGLGCGAGAAAAAG-----AGCGGLGCGGAGGAGAG--AAAAAAGAG 380
 DB 730 TSSGCGAG 789
 381 GVRGCGGGLGSGAGAGCGGAGAGAAAAAGAGAGCGGCGGLGSGAGAG----- 425
 DB 790 G--DGGNGGDDGFFGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
 426 GAGGGLGCGGAGAAAAAG-----GCGGTGGCG-----GAGAGAGAGAG 462

DB 848 CTGCGNGTGTGTGAG 899

RESULT 8
 ID 044358 PRELIMINARY; PRT: 871 AA.
 AC 044358;
 DT 01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
 DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
 DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
 GN FLAG.
 OS Nephila clavipes (Orb spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorpha; Entelegynae; Araneidae; Tetragnathidae; Nephila.
 NCBI_TaxId=6915;
 RX MEDLINE=98153262; PubMed=9480768;
 RA Hayashi C.Y., Lewis R.V.;
 RT "Evidence from flagelliform silk cDNA for the structural basis of
 RT elasticity and modular nature of spider silks."
 RL J. Mol. Biol. 275:773-784(1998).
 DR EMBL: AF027972; AAC38846.1; .
 DR InterPro: IPR000087; .
 DR InterPro: IPR000209; .
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
 FT NON_TER 871
 SO SEQUENCE 871 AA; 71039 MW; 1FA1EB37E0C5983A CRC64;

Query Match 34.1%; Score 921; DB 5; Length 871;
 Best Local Similarity 41.0%; Pred. No. 5.7e-47;
 Matches 245; Conservative 22; Mismatches 212; Indels 118; Gaps 22;

19 GGLGCGAGAGAGAAAAAGAGCGGCGY-----GLGSGAGAGCGGAGAAAAAGAG 71
 DB 245 GCGAGPGAG--GAYGPGGVTGTGAGGPGGAGPGGAGPGGAGPGGAGPGGAGPGG 303
 72 GCGGCGGGLGSGAGAGAGAGAAAAAGAGCGGCGGLGSGAGAG--GAGAAAAAGAGAG 130
 DB 304 GCGAGPGGCTGPGGAGPGGCTGPGGAG-----PGGTGPGGTGPGGTGPGGTGPGGAG 359
 131 GY--GGLGSGAGAGCGGCGGCGGAGAAAAAGAGAGCGGCGY-----GLGSGAG 174
 DB 360 GYCPGAGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAG 419
 175 RGGLGCGAGAGAAAAAG--GAGAGCGGCG--GLGCGAGAGCGGTGGLGSGAGAGRG 228
 DB 420 GAGLGGAGCGAGTSGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479
 229 GAGAAAAAGAG--GCGGLGAG-----GAGAAAAAGAGAG 261
 DB 480 GARGAG 539
 262 G-----GYGGLGSGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
 DB 540 GNVGPGSGGCGGVPGSGGCGGVPGSGGCGGVPGSGGCGGVPGSGGCGGVPGSGG 529
 305 GGLGSGAGAGRGGLGAG-----GAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 354
 DB 600 GGVGTGVPGAGAGGPGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
 355 -----GCGGLGAG-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 397
 DB 660 GCPYCGAG 717
 398 GCGAG 455
 DB 718 GGTGPG--GAGPGGCTGPGGAG--GCTGPGGSGGCGGPGGSGGCGGPGGAGGTG 774
 456 SAASAAASRLSPQASSRVSAVSNLVASGPTNSAALSTISNVVSIQAGNPLGSG 512

```

Db 775 SESGCGPBGSGPGSGPBGSGPG---GSGPGGCGYCPGSGSPSSFPV---GSGCPGSG 826
RESULT 9
ID 053559 PRELIMINARY; PRT: 1489 AA.
AC 053559;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE PGSS-FAMILY PROTEIN.
GN RV3514 OR MT023.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Kiroh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter K., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
EMBL: AL022022; CAAL1751.1; -.
DR Tuberculist; RV3514; -.
DR InterPro: IPR000084; -.
DR InterPro: IPR002173; -.
DR Pfam; PF00934; PE; 1.
DR PRODOM; PD001223; -.
DR PROSITE; PS00583; PKK_KINASES_1; UNKNOWN_2.
SO SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3CBAF3A CRC64;

Query Match 33.8%; Score 912.5; DB 2; Length 1489;
Best Local Similarity 39.1%; Pred. No. 2.9e-46;
Matches 263; Conservative 19; Mismatches 202; Indels 189; Gaps 34;

OY 2 ASMGCGQMGRIIRIRYGVGLGSGAGQ---GAGAAAAAGAGGCGGGLGSGAGRGGO 59
Db 178 AGGAGGWLFG---VGGAGGAGGAGGCTGTGAGGPGGLTWGGGAGGAGGAGGTGTGAGGRA 234
OY 60 ---GAGAAAAA---GG---AGGCGYGLGSGQ---AGRGGLGSGQ---GAGAAAAAGCV 103
Db 225 ELRLGAGAGAGAGGTGTGATGTGCGHGVGGDGMWLAPGAGGAGGAGGAGAGSDGA 294
OY 104 ---GGCGGLGCGAGGAGCA---AAAAAGAGGCGYGLGSGQ---AGRGSGSGQ--- 148
Db 295 LGGTGCTGCTGTGAGAGGAGRALLLGAGGCGGGLGAGCGGTGTGAGGAGGAGGAGG 354
OY 149 ---GAGAAA---AAAGGAGGCGYGLGSGAGRGGLGSGGAGAAAAAGAGAG 195
Db 335 GVGGAAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 413
OY 196 ---GCGYGLGSGQ---AGCGYGLGSGQ---AGCGYGLGSGQ---AGCGYGLGSGQ--- 231
Db 414 FAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 473
OY 232 AAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 275
Db 474 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 533
OY 276 GQ---GAGAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 328
Db 534 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 592

```

```

OY 329 AAAGAGCGGGLGSGQ---GAGAAAAAGAGGAGGGLGSGQ-----AGGAGA 370
Db 593 GAGGAAGTGTGCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 652
OY 371 AAAAAAAGAGV-----ROGTYGLGSGQA-----GRCGAGAGAAAAAG 410
Db 653 GKAGGSSAGGAGTSSGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 712
OY 411 G-AGCGTGTG---GGCGV-----GAGGLGSGQ---GAGAAAAAGA----- 443
Db 713 GAAGTGTGTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 772
OY 444 ---GCGYGVGSSGASAAASRLSSPOASSRVSSAVSNLVASSPTNSALSTISNVVS 501
Db 773 ICGTGCGDGTGGAAGAGAGG-----AACTGTGTGAGTGTGAGGAGGAGGAGGAG 810
OY 502 QIGASNPGLSGCD 514
Db 811 ---GVGAGAGGCGGD 821

RESULT 10
ID 09NHM4 PRELIMINARY; PRT: 2249 AA.
AC 09NHM4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GN FLAG.
OS Nephila clavipes (orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
CX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20156766; PubMed=10688794;
RA Hayashi C.Y., Lewis R.V.;
RT "Molecular architecture and evolution of a modular spider silk protein
RT gene.";
RL Science 287:1477-1479(2000).
EMBL: AF218621; AAF36090.1; -.
DR InterPro: IPR000087; -.
DR InterPro: IPR000209; -.
DR PROSITE; PS00138; SUBRTILASE_SER; UNKNOWN_1.
FT NON_TER 1
FT TER 2249
SO SEQUENCE 2249 AA; 174871 MW; 88C8B168A147CDA0 CRC64;

Query Match 33.3%; Score 900; DB 5; Length 2249;
Best Local Similarity 36.0%; Pred. No. 2.2e-45;
Matches 245; Conservative 27; Mismatches 220; Indels 188; Gaps 24;

OY 6 GGGQMGRIIRIRY---GSLGAGAGGAGAAAAAGAGGCGY-----GGLSGAGR 56
Db 224 GGVGPGGAGPGGYPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGL 283
OY 57 GG-----QGAGAAAAAGAGAGGAGG---GGLSGAGRGGL----- 89
Db 284 GGAAGPGAGLGGAGPGGAGTSGAGPGGAGPGGAGGAGGAGGAGGAGGAGGAGGAGGAG 343
OY 90 GGGGAGAAAAAGV---GCGGLGCGGAGQ-----GAGAAAAA 123
Db 344 GAGAGRGAGRGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 403
OY 124 AGAGAGQ-----GYGLGSGAGRGSGGAGAAAAAGAGAGGCGY-----GGLG 169
Db 404 PGAGPGNVGPGRSRPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 463
OY 170 SQAGRGGLGSGQ---GAGAAAAAAGS-----AGCGYGLGSGQA----- 207

```

OY	2	ASMTGCGQMGRIIRITRGVIGLGGSGAGCAGAAATAAAGAGCGGV-----GGLGSOGA	54
Db	149	AGOSGN-----GGSAGLWNGGAGAGGSGGGAAGGNGNGLFGAGTGTGIGTTGA	201
OY	. 55	-----GRGGGCGAGAATAAAGAGGAGGGYGGLGSSQGARGGLGEGOGAGAA--AAAGV-CQG	106
Db	202	PGMAMGTGTGNG- GKNALLIIGGGGLGAGAGMGTGGTGTTGTGNGNGNGLLIIGAGVGAGG	260
OY	107	GLGGGAGGAGGAAAAAACG-----ACGGYGGLGSOGA-----GRGSGG	147
Db	261	GIGGGGTGAGGAAGAGGTGNGAGGLFMNMGDGDGAGGCGGDBGAADAAASAGTGGKGG	320
OY	148	QGAGAAAAAAGA-----GQGGYGGLGSOG-----ACRGLGGQC--A	183
Db	321	QGGDGCTGAGGAGFVLPHGHGAGAGMGQGGGTGNGAGAGDDTTVIAAGTGGEGTGGAA	380
OY	184	GAATAAAAAGA-----GQGGYGGLGSOGA-----GQGGY	212
Db	381	GAGCAAGARCALLISGLLAGVAGAGTGGTGTGGAGADAAAVYGCANDBPGFAAGKGN	440
OY	213	GGLGSQG-----AGRGLGGQG-----AGAAAAAGS-AGQGLGQG-GAGAAAA	254
Db	441	GGIGAAYTVGVAQDGGTGGKGGTGGAGAGNDASTGNPCKGGDGGIGAGAGAGAAAG	500
OY	255	A-----AGGAGQGGYGGIGLSOGA-GRG-----QGAGAAAAA	286
Db	501	TGNGCHAANTGPBGGGGTGTCGNCNGCTGGVNCADNTLMPTPGAGEBPQAGAGAGAGGA	560
OY	287	GG-----AGQGGYGQGG-----AQGGYGGLGSQG-----AGRGLGQ-----GAGA	324
Db	561	GGPGGTGGTGGGNGGAGNGGNGNGNGNAGNNSTNAFYBGBGAGGTGGAGGAGG	620
OY	325	AAAAAAGAGAGQGGGLGSG-----AGAA	347
Db	621	AANGSTAGSQGTGGVGGDGGAGAGNGGGKACTGNSGNFGVDGEAFSGAGANGVGAA	680
OY	348	AAAAGAGQGGGLGSGAGAGAA-----AAAAAAGCVROGYGGIGLSOGA	395
Db	681	GANGTGGSGGNGDGGAGGTGGAGNGIPGTGTPEAGGTGAKGDDGGDGGAGAGGNAAG	740
OY	396	GRGGGAGAAAAAAGAG-----QGTCGILGGGQGVGAGLGGGGAGAA	438
Db	741	GAGGGGAGNAGCGAGAGNAVTPEDGYEKAPHDHAGSGGGGGKNG-QQGGSGGTG-SGG	798
OY	439	AAVGAQGGGCGVGV-----SCASAASAAARSPPQASRVSAVS--NLVASCPPTNSA	491
Db	799	APIGGGAGTGGSGGHACKGAGGICGACTTTTYGNGNGMDGNGGNAAGAGNGSSOD	858
OY	492	LSTTSNVYSQTGAS--NPGLSG	512
Db	859	FQGNNTSGASGSGGNGNAGTAG	881
RESULT	12		
ID	002402	PRELIMINARY;	PRT; 738 AA.
AC	002402;		
DT	01-JUL-1997	(TREMBLrel. 04, Created)	
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)	
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)	
DE	INSOLUBLE PROTEIN.		
OS	Plinctoda fucata.		
OC	Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidae;		
OC	Pterioidae; Pterizidae; Plinctoda.		
OX	NCBI_TaxID=50426;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=97320490; PubMed=9177341;		
RA	Sudo S., Fujikawa T., Nakakura T., Ohkubo T., Sakaguchi K., Tanaka M.,		
RA	Nakashima K., Takahashi T.;		
RT	"Structures of mollusc shell framework proteins.";		
RL	Nature 387:563-564(1997).		


```

Query Match 32.5%: Score 877.5: DB 5; Length 1468;
Best Local Similarity 36.1%: Pred. NO. 3.2e-44;
Matches 262; Conservative 51; Mismatches 182; Indels 231; Gaps 28;

OY 2 ASMTGQOGRIRIRKGYGLGCGGAGGAGAAAAAGAGGCGGYGLGSG-----AGR 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 SASRSGAGLGCGAAGAGSLGGLG---GAGSAAASA-CAGLGGIGAGSGSSSSAAS 412
OY 57 GGGGAG-----AAAAAGAGGG-----YGLGSGQ----- 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 GASGAGEYIVIDRRSSAASAAASGAGSLGGLGGLGCPYGTGILNGVSSAALGAGLG 472
OY 84 -----AGRGGLGGGAG-----AAAAGYQGGLGCGAG-----QG 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 473 GVTGAGAGSLGGLGAGYSAVGPAGAGLGGVAGGSSGSSAASARSAPYVIEDG 532
OY 117 AGAAAAAGGAGGCGGYGL-----GSGGRCGSGGQGA 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 533 SSASAAAAAGSAGSLGGLGAGWPLGIGIPNEYSSASATSAAGSTCAGLGGSSAAGS 592
OY 151 GAAAAAG-----GAGGCGYGL----- 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 593 SAASASGAGAPYIVIEDGSSAASAAAGSGAGSLGGLGAGMTLGLGIPNGVSSAS 652
OY 169 -----GSGGRCGLGGG-----AGA-----AAAAAGAGGCGYVG 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 653 ATGSAAGSTGGGLCGSGAAGSSLASAGAGAPYVIEDGSSAASAAASGASGVGG 712
OY 202 LGGGAGGCGYCGLGSGAGRGGLGCGAGAAAAAGAGGGLGCGG---AGAAAAAGG 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 713 L-LGGA-LGPLGGIGPLGASAGAGLGGVGAAGTSLGIGIGVASTAGSAGAGLGG 770
OY 259 AGGCGYGLGSGGAGRGCGAG-----AAAAAGCAGCGYGCAGC-QGYYGL 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 IGAGSSGSSSAAASAGSAGAGEYIVIDRRSSAASAAASGASGPGGIGLGVWPLGCI 830
OY 308 GSGGRCRGGLGCGAGAAAAAACGAGCGGCGAGAAAAAGCAGCGGCGGAGG 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 831 GPLGSSASASGAGLG---GYGAATGSLGGIG--GAGSAAAGSAGAGLGGAGSSGSS 885
OY 368 AGAAAA-----AAAAAGYRGGYGLGSLGSGAG-RCGQGA-CAAAAA 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 886 SVASAASTSGAGEYIVIDRRSSAASAAAGSGAGSLGGLGGLGIPYGTGILNGVSSAS 945
OY 409 AGGAGGCTGGLGCGGAGLGGGAGAAAAVAGGCGYGVGS-GASASAAASRLSS 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 946 ALGAGLGGVGYACASGL--CGLGGCGASAAAGSAGAGLGGVAGGSGFGSSAAASGASGA 1003
OY 468 PQA-----SSRVSSAVSNLVASG-----TNSALSLPISNVVSLGAS 506
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1004 GEYIVIDRRSSAASAAAGSGGPGGLGGLGVWPLGIGIPNGVSSASATGSAAGSTGA- 1062
OY 507 NPGLSG 512
   : : :
Db 1063 --GLGG 1066

```

Query Match Similarity	32.2%	Score 868.5	DB 2	Length 1660
Best Local Similarity	41.5%	Pred No. 1.2e-43		
Matches	245	Conservative	22	Mismatches 175
			Indels 149	Gaps 30

QY	17	GYGGLGGGAGGAGAAAAAAGAGGCGGTGGLGSG---	GAGRGGGAGAAAAAAGAG	72
		:		
Db	649	GKGGAGGAGATGTCATGATVTHSGNGGKGNGADATV	TAAGANCGKAGGNGGLTGGDG	708
QY	73	GGGTGGLSGGCA-----	GRGLGGGAGCAAAAAGC----	102
		:		
Db	709	AGGDDGSGGAAGANGANVGEDGADGTLGCGPCG	SEANGGCGGTGGGAGAGAGDGGGAGSS	768
QY	103	-VGGGGLGGG--AGGGAGCAAAAAGCA-----	GCGGTGGLGSGGCA--GRRGSG	146
		:		
Db	769	ALGSGGNGRGGAGAGGAGGAGAGGAGSVSGD	CGPKGAGGAGGAGAGSGGGGAGK	828
QY	147	GCGAGAAAAAAGAGGAGGCGTGLSGGAGRGVGLG	GGGAGAAAAAAGAGGAGGCGTGLSGG	206
		:		
Db	829	ASGADSAEAVGAGGKGGGTGVTG--GGGGRGGG	-GAGGAAPAGQVSHGVTGGTGG	886
QY	207	-----AGCGTGTGLSGGCA-----	GRGLGCG-----GAGGCA	243
		:		
Db	887	GLGAGAGCGGCGDGHGSDGDDGCGDPCGAG	GLGGLGCGSGNCTRAASGYDASDHGPGSSG	946
QY	244	LGGGGCA-GAAAAAAGCA-GCGGTGG-----	LGSGGCA-GRGGGAGAAAAAAGCA-----	289
		:		
Db	947	NGGNGNGAGQASVAGGAGGNGGDDGAGRGVGG	CGAGNGGDDGAGAGNGANSQPGSDATA	1006
QY	290	-GG--GGTGGG-----	AGCGGTGGLGSGGAGRGGLGG--GAGAAAAAAGAGGCGG	338
		:		
Db	1007	LGPGGNGGGGAGGAGGAGGAGGAGGAGSVSG	DDGAGGAGGAGGNGGCVGASGSGGANG	1066
QY	339	LGG-----GAGAAAAAAGCAGCGGTGCGG	CAGGAGCAAAAAAAGCA-----GCVRGGTGTG	391
		:		

Db 1067 ANGIDSTIGTGAGGCGGDDGAGGVGCHGDPGVGAAPSGTVGSHGTGV--GGDGLG 1124
OY 392 SQG--AGRGCQ-----GAGAAAAAGAGOGGTGLG-----OGV----- 425
Db 1125 GAGGVGAGGNGGIGITVCGAGAGNGCDPAGRGGLGSDSGNGTSAANGVDASKHGP 1184
OY 426 ---GAGLGGGAGAAAAVG-AGGCGTGG-----VSSGASASASAAA 462
Db 1185 LTGGDGVGVGNKAKAAAAAGDGGGDNAGLFGDGGAGGDADGTAAEA 1235

Search completed: July 3, 2001, 15:05:41
Job time: 606 sec